

Star wars threat to ABM treaty

The Pentagon recipe for testing star wars technology without technically violating the anti-ballistic missile treaty is an artificial exercise likely to do more harm than good.

THE Jesuits earned themselves a bad name in the sixteenth century not only for their enthusiastic conduct of the inquisition but for their casuistry, their willingness to bend general rules to the circumstances of particular cases. The Jesuits are now much reformed, but the US Department of Defense is well on the way to succeeding to their odious reputation. The Pentagon's publication last week of its recipe for testing the viability of its star wars plans without violating the Anti-Ballistic Missile (ABM) Treaty with the Soviet Union (see p.3) is as fine an example of the jesuitical arts as there has been for years. In defence of its reputation, the department may say that it would never have dreamed of producing a document certain to give as many hostages to fortune if Congress had not required it to do so. But the Congress, whose love for star wars has yet to blossom, intended to offer Mr Caspar Weinberger, the US Secretary of Defense, a poisoned chalice; it would have been possible and prudent for him to have replied in terms less obviously calculated to undermine the ABM treaty.

What the Pentagon has done is to exploit the verbal inconsistencies of the treaty, as drawn in 1972 when the two signatories had no immediate cause to contemplate anti-ballistic missiles of other than the then conventional kinds. The chief interdiction (Article 1) is that neither side should deploy an ABM system except at the sites agreed (originally two, reduced to one in 1974). An ABM system is defined (Article 2) as one "currently consisting" (in 1972) of interceptor missiles, launchers for them and radars designed for use against the ballistic missiles rather than, for example, in air defence. The treaty says that none of these components should be deployed (except as specified) and makes plain that the interdiction applies not only to operational components but to components "undergoing testing". Elsewhere (Article 5), the treaty goes on to say that the signatories shall not "develop, test or deploy" ABM systems or their components which are based in space, in the atmosphere, on the sea or which are even movable on dry land — a point emphasized by the way in which the US delegation to the negotiations asked for, and got (in January 1972), an assurance that it would be a violation of the treaty if either party built a radar or a launcher for an intercepting vehicle that was not fixed on the surface of the Earth.

Violation

The bearing of these provisions on the way in which two parties to the treaty conduct themselves ten years later is ambiguous only to those who wish that it should be so. Most people will, for example, understand the string of verbs "develop, test or deploy" to indicate that development is the stage in a technical process that precedes testing, which is in turn a preliminary to deployment. But on that reading, the Strategic Defense Initiative will violate the ABM treaty once the feasibility of space-based missile defences has been investigated (by research) and the development of hardware has begun. The distinction between research and development is notoriously arbitrary, in civil as well as military industry, but because the star wars project will be increasingly concerned with demonstration projects (aimed at convincing the military, members of the Congress and, no doubt, US taxpayers) that the game is worth the candle, it is inevitable that the United States will soon be accused of having broken the treaty. "But it's only research", will be the cry, which will be

defensible only on the hollow grounds that the US Congress has not by then agreed to provide funds for more than the research programme as now defined.

Sham

The notion that it will be possible to test the efficacy of novel kinds of missile interceptors without violating the treaty by pointing them at satellites rather than at ballistic missiles, or that tracking radars will be permitted if operated at reduced power, is similarly a dangerous sham. Although a successful demonstration that this or that new kind of ray gun can destroy a satellite will not, for example, also test the effectiveness of the target-tracking systems that would be needed for an operating defence against ballistic missiles, a few experiments like these could provide the knowledge on which full-scale missile interceptors based in space may be confidently designed. It should be even easier to extrapolate from observations with low-power radars or infrared sensors to the components needed for a defensive system. So there is every reason why the course charted in last week's document will provoke the Soviet Union to fierce protests. The Pentagon should also worry about the extent to which this advance advertisement of how it will break the spirit if not the letter of the treaty will generally undermine even the support of those who accept that the case for a programme of research on anti-ballistic missile defence can be sustained.

But how else to arrange that the feasibility of star wars can be tested? As it happens, the ABM treaty provides for a mechanism for doing just this — the creation of a Standing Consultative Commission to consider questions arising under or about the treaty, including (Article 13(d)) "possible changes in the strategic situation which have a bearing on the provisions of this treaty". By agreement, the commission was lumped together with that required by the first SALT agreement. This, no doubt, is the body to which the United States will turn if and when the time comes to raise with the Soviet Union the case for deploying some kind of system based on star wars. (This is also the forum in which the United States should have raised its proper doubts about the function of the Soviet radar, now being built near Krasnoyarsk in central Siberia, which is of a type whose siting there is forbidden by the ABM treaty.) It would be better to bite that bullet sooner rather than later, if only to keep some kind of relic from the days when arms control agreements seemed desirable and feasible.

But why bother? Nobody forgets that part of President Reagan's objective in embarking on star wars was his wish to make arms control agreements unnecessary. If the two superpowers (or even just one of them) could shelter behind solid missile defences, those endless visits to Geneva might be dispensed with. This argument (some say wish-fulfilment) is often heard in the United States, much less commonly in Western Europe. There are two difficulties. First, the United States is already engaged in arms control negotiations at Geneva (adjourned until later in the month). Second, the arguments that made the ABM treaty seem worthwhile in the 1970s are still valid even in the changed circumstances of the 1980s. The objective then was to protect the strategic retaliatory forces of each side from being rendered ineffectual. The proliferation of warheads and their increased accuracy has since made counter-force strategies



credible, but only just. But there is no reason to believe that a working star wars strategy is more easily realized. In short, and for as long as star wars remains a hypothesis to be tested, the need for arms control is as pervasive and as irresistible as President Reagan found during the first few months of his first term, in 1980.

But what is to be done about Soviet duplicity in this and other connections? At times, the US administration sounds a little like the lawyer whose defence of his client against a charge of murder takes the form of asserting that the client did not kill the victim and, then, that he did so in self-defence. The Krasnoyarsk radar is a sign (in the absence of argument to the contrary) that the Soviet Union has not been resting on its laurels, the anti-ballistic missile system around Moscow. It may also be the case that the Soviet Union is already hard at work on a research programme not very different in character from star wars itself, as the US administration claims, but that is irrelevant. The lesson that needs now to be learned is that, for as long as these schemes remain dreams (not necessarily illusions) they should not be allowed to affect policy. And if they are found feasible, their deployment will have to be discussed between the superpowers. Why not accept that now, not later? □

Fun made dull?

British school mathematics may be made worse by the government's recipe for reform.

WHAT should a young person know of mathematics? And how should mathematics be taught at school? From time to time, these questions recur almost everywhere, as some mathematicians would say. Under the idiosyncratic guidance of Sir Keith Joseph, Secretary of State for Education and Science, the British government has just embarked on yet another attempt to breathe new life into a jaded curriculum. Things may seem to have changed a lot (but not for the worse) since earlier decades, when British governments behaved as if they had no right to tell the local education authorities, which administer British schools, what should be taught, but only a kind of admonitory right to complain if teaching and (more often) other aspects of school life were hopelessly awry. But on this occasion, the government has decided to stiffen its exhortation to reform by the appointment of more than 300 full-time teachers whose sole responsibility will be to ensure that the pace of reform of mathematics teaching does not flag. Those with long memories should be forgiven for reminiscing about the early 1960s, when the government encouraged renewal of the mathematics curriculum, and when local education authorities showed willing by appointing people called "advisers" to help teachers change their teaching habits.

The objective now is also much what it was two decades ago, when the general opinion was that mathematics should be made to seem more like fun to students, that it should seem more "relevant" and that teaching should concentrate more on conceptual matters than on the learning of manipulative skills. (In Britain in the 1960s, although concepts such as that of sets made their way into the curriculum even of younger children, the "new math." never made quite the headway that terrified parents and some younger people in the United States.)

The occasion for the British government's new venture is a booklet by Her Majesty's Inspectors (of schools), called *Mathematics from 7 to 16* (HMSO, £2.00), which argues much the 1960s case for reform, but in a way that has predictably excited the British popular press. The most inflammatory passage is that which argues cogently that young people should be helped to use calculators as a means of carrying out calculations, not simply be allowed to use them for confirming the results of calculations done by other means, and which goes on to say that "long division... should no longer be generally taught" and that "logarithms, as aids to calculation, are obviously redundant". The rumble of conservative protest has been palpable, but even radicals have been taken aback.

Conservatives should not be afraid that the end of their world

has come. The new document is a recipe not so much for upheaval as for consolidation, even for consolidation about modest goals. This is not surprising, since the school inspectors have themselves relied on a study of school mathematics carried out by a government committee under Dr (now Sir) William Cockcroft and published three years ago. The inspectors echo that earlier plea that the mathematics taught in schools should consist of a central core of knowledge and understanding determined by the needs, interests and abilities of most students. On this view, mathematics at school is largely a utilitarian study. Even the section in the manifesto on "conceptual structures" is largely devoted to a sensible discussion of the reasons why students should appreciate the importance of estimating the results of calculations and approximating to them (or even to the insoluble problems of the real world). The document repeatedly urges that teachers should engage their students in practical work, but it is partly redeemed by its advocacy of open-ended questions in mathematics examinations.

The more serious objection to what is now proposed is its dullness, or its potential dullness in the hands and heads of many teachers. While the language of sets may, in the 1960s, have seemed a licence to some teachers to make algebraists of five-year-olds, the repeated exhortation to teachers now to bear in mind the practical needs of their students, especially to find jobs, will give many teachers the idea that mathematics should be turned into sophisticated mensuration. And while the inspectors' document repeats in humdrum language the belief that mathematics is more a part of modern life than of that of any previous age, it does very little to suggest to teachers that they should help their students understand how mathematics has become so powerful. Thus, absurdly, in spite of the inspectors' advocacy of computers as a means of making calculations, nothing is said about the benefits of understanding that even the simplest techniques for arithmetical manipulation (not to mention long division) involve the use of algorithms of a kind that also underlie computer programs. Fortunately, most teachers and many of their students already know better.

All is not lost, however, for the inspectors' prospectus is described as a discussion document, on which opinions will be welcome. Here are some complaints that the inspectors (and their minister) might usefully accommodate. First, because mathematics is as important as the document says, students of all ages and ability should be helped to understand how these circumstances have arisen (which argues for a little history) and why mathematics is so powerful (which argues for school-leavers knowing something of what can be accomplished by the calculus even if they have not learned to differentiate a simple function). Second, and consistently with the inspectors' utilitarian cast of mind, the curriculum should give students a general understanding of how problems of the real work are turned into problems in mathematics which, in practice, are soluble only exceptionally. Third, students need to have their spirits lifted by repeated demonstrations of the sheer neatness of much of mathematics, whence much of its power. But they need also to appreciate that, like other parts of science, mathematics remains unfinished business, as can be told at the earliest ages from the lack of an analytical formula specifying the prime numbers.

Can Sir Keith Joseph's band of advisers hope to accomplish all this? With luck, they will do much good, but there are two respects in which they will be handicapped from the outset. First, the mathematics teachers in British schools are too few and are also, too often, as unconfident of the material they hope to teach as are their students. The immediate remedy is for more in-service training, as the government agrees. The long-term solution may be unattainable, given the demand for people who can count in the economy at large. Second, a point which the government does not so far concede, there is a need for yet another deliberate attempt to redesign the mathematics curriculum by practical work by teachers working in the schools, not by edict from on high. Plainly, as things have turned out, what British school mathematics most needs is a flood of protest to the Senior Chief Inspector, Mr E. J. Bolton. □

Star wars

DoD makes a monkey of anti-missile treaty

Washington

THE Pentagon's Strategic Defence Initiative (star wars) office has produced the first detailed account of how it intends to carry out field tests of new ballistic-missile-defence hardware over the next five years without violating the 1972 Anti-Ballistic Missile (ABM) treaty. The explanation, contained in a report ordered by Congress when it approved the star wars budget last year, is likely only to inflame criticism that the Reagan administration is playing fast and loose with the treaty requirements.

Most of the field tests would be justified on the grounds that they do not involve "ABM components" or "prototypes of ABM components". The treaty does not prohibit the development of weapons to shoot down satellites, cruise missiles or tactical ballistic missiles. So when railguns and rocket-propelled projectiles are tested in Earth orbit in the early 1990s, their targets will be satellites, not missiles. The Department of Defense (DoD) report says that such tests will demonstrate the technology needed to defend space-based ABM battlestations against Soviet anti-satellite weapons. But these technologies are also candidates for ABM systems proper, although DoD says only that the tests will "permit a decision to be made on the applicability of more advanced technology for ABM purposes".

Tests of lasers and space-based tracking systems, on the other hand, would be justified on the grounds that they will have only "limited capability". DoD says because the hardware to be tested is not advanced enough to operate in a fully fledged ABM system, it is not a "prototype". At least some of the limitations may be artificial. Thus a system to track Inter-Continental Ballistic Missiles (ICBMs) will not qualify as an ABM component prototype because it "may measure the signature of booster plumes, but not in real time".

Congressional critics of Star Wars were quick to point out, after the report was issued last week, that the Pentagon's reliance on the anti-satellite-weapon loophole is further evidence of the Reagan administration's opposition to any progress on a treaty limiting testing of anti-satellite weapons. Representative George Brown (Democrat, California) has called for a renewal of a congressional ban on further anti-satellite testing. He also called plans to test star wars hardware against satellites a "circumvention of the spirit of the ABM treaty".

An analysis by legal scholar Abram Chayes of Harvard University reaches a similar conclusion. He says that attempts to develop star wars technology under the

label of anti-satellite or anti-tactical-ballistic-missile programmes are "legally disingenuous" as well as technically costly. Chayes points out that the course charted by the Reagan administration will inevitably generate disputes over its legality under the treaty and argues that, instead of being simply a "research" programme that will leave open a decision to renounce the treaty in the mid-1990s and proceed with deployment, star wars as now conceived may make that decision a foregone conclusion.

Star wars

Academics sign on for research

Washington

WHATEVER their first reservations, US universities seem to be having no scruples about cashing in on President Reagan's star wars programme of ballistic missile defence, known formally as the Strategic Defence Initiative (SDI). The SDI organization's innovative science and technology office, which will spend 5 per cent of the star wars budget, has started to award multi-million dollar contracts to university/industry consortia for basic research with possible applications to missile defences. The office, established only last November, will soon have committed over \$50 million for basic research. More than 50 US institutions will be working on basic star wars research within the next few months.

The three consortia already named by the innovative science office will be working on non-nuclear power sources suitable for space-based systems (\$19 million over four years), optical signal processing (\$9 million over three years) and advanced composite materials (\$15 million over three years).

Smaller institutions are not being left out. Besides such well-known research establishments as the California Institute of Technology and Massachusetts Institute of Technology, the consortia include, for example, the University of Texas at Arlington and the University of Alabama at Huntsville.

Applied (as distinct from basic) research, which accounts for the greater part of the SDI budget, is being carried out in-house in government laboratories or in industrial companies; several hundred SDI research contracts have already been negotiated. SDI has a total budget of \$26,000 million over the five years from 1 October 1984.

The director of the innovative science and technology office, James Ionson, says he has no illusions about the magnitude of the technical problems confronting SDI, but he is convinced that solutions to many

The DoD report suggests two other ways to get around the provisions of the ABM treaty. The treaty specifically allows field testing of fixed land-based ABM systems or components. This was the administration's justification for the experiment last year in which an interceptor launched from Kwajalein Missile Range in the Pacific hit an ICBM launched from White Sands Missile Range in California. Several further tests of both radars and interceptors is planned under this provision.

The report also hints at a more drastic way to evade the treaty's constraints. Noting alleged Soviet violations of the treaty, the report says "we do reserve the right to respond to those violations in appropriate ways, some of which may eventually bear on the Treaty constraints as they apply to the United States".

Stephen Budiansky

of them will eventually emerge from speculative basic research of the sort supported by his office. According to Dr Ionson, many ideas that at first "seemed crazy" have now demonstrated their potential and will be included in the SDI effort.

All the basic research supported at present through Ionson's office is unclassified, but the classification requirements for SDI research are under scrutiny by the Department of Defense (DoD). Ionson is confident, however, that DoD's policy of relying on classification alone to control on-campus basic research will be maintained.

The recent disruption caused by DoD at a meeting of the Society of Photo-optical Instrumentation Engineers (see *Nature* 18 April, p.569) raised the possibility that DoD might in future make more use of Export Administration Regulations to restrict the flow of sensitive technology abroad. According to Dr Ionson, use of such regulations is most likely when commercial interests become involved but each case will be considered on its own merits.

Export regulations will undoubtedly be the subject of politically-delicate negotiations if overseas institutions participate in star wars research, because foreign governments will want to ensure that they receive the benefits of technology spin-off. But Defense Secretary Caspar Weinberger has still not given the formal go-ahead for foreign collaboration.

Among the major research threads supported through Dr Ionson's office is optical systolic computing, which could bring within reach the enormous computing power that would be needed for a star wars defence. Ionson is also looking at the possibilities for molecularly engineered dielectrics for capacitors, plasmoid weapons and the protection of satellites in radioactive environments by pumping out the Van Allen radiation belt.

Tim Beardsley

Japanese Science Council

Elected body goes out fighting

Tokyo

THE Science Council of Japan (JSC)'s final meeting as a directly elected body ended last week in the liberal tradition it has long upheld. One of its final actions was to call on the government to improve the status of female scientists.

Since 1948, Japan's unique "Parliament of Scientists" has provided a means for Japanese research workers to express their opinions to government. Every three years, 210 representatives from every branch of science have been elected to the council by a direct vote of more than 200,000 of Japan's scientists. Committees organized by the council have been instrumental in shaping the grants policy of the Ministry of Education, Culture and Science (MESC), in pressing for the establishment of new institutes in areas where they were thought most needed and in upholding scientists' rights. The president of JSC is automatically a member of the Council for Science and Technology, the highest science policy-making body in the land, and thus has the ear of the Prime Minister.

In recent years, however, things have not always gone smoothly between the government and JSC. As MESC relied more and more on advice from its own internal science council, JSC came to see its influence on grant policy decline. In turn, voters became apathetic and, as the government sees it, the council came to be influenced more by those with a "political" axe to grind than by those concerned with purely scientific matters. Apparent opposition to nuclear power by JSC particularly irritated the government.

In 1982, the government, which pays all JSC's expenses, finally demanded reform, and when it proved hard to agree, reform was forced on JSC. The result is that from now on, members will no longer be directly elected but will instead be appointed through the scientific societies (see *Nature* 305, 361; 1983).

The call to improve the lot of women scientists stems from surveys carried out by JSC over the past few years of the special difficulties encountered by women scientists. It was plain that few women were reaching the top posts: there are no female full professors in the science faculties of the principal universities, few women among the laboratory chiefs of government research institutes and a relatively small percentage of women entrants to science courses in top universities.

These observations mirror problems for women in Japanese society as a whole. The average salaries of women are barely half those of men, not so much because of different rates of pay for the same job but because women generally occupy lower positions. Surveys show that three-quarters of companies hiring university graduates (who will eventually go into management posi-

tions) hire only men. Of those that do hire women, half will not allow them any possibility of promotion and two-thirds will not give them access to training or education programmes. "Retirement" at marriage is a not unusual condition of work.

JSC's subcommittee on the status of female researchers carried out its own survey on the position of women scientists after an earlier plea, in 1977, failed to persuade the government to conduct a proper survey. Some 2,000 women scientists and 1,400 male scientists were asked to complete questionnaires, and the results confirm women's poor promotion prospects.

Taking into account academic record, the chances of a woman becoming a full professor appear much lower than those for a man. Most women who successfully complete their doctorates cannot hope to gain a post higher than assistant lecturer and the higher they go, the more difficult further promotion becomes. Professor Katsuko Saruhashi, one of the few women to reach laboratory director level (within the govern-

ment's Meteorological Research Institute) and a veteran campaigner for the rights of women scientists, says it is bad for the nation not to make use of women's abilities and that what is now being demanded is a proper government survey that could be used to make the reforms that would improve the position of women.

On the government's recent record, however, it is unlikely that her wishes will be fulfilled. The Diet will vote next month on an equal opportunities bill intended to give the government an opportunity to ratify the United Nations Convention on the Elimination of All Forms of Discrimination Against Women. The bill has, however, been thoroughly emasculated by employers' representatives, and apart from abolishing some old restrictions, such as those that prevent women working late at night, it provides little in the way of new rights. There are so many loopholes in the bill that it would be almost impossible for legal action to be taken on its provisions. Although it calls for women to be given equal treatment to men in job recruitment, allocations and promotion, there are no substantial penalties for non-compliance.

Alun Anderson

Romania

Nuclear power in the distance

PRESIDENT Nicolae Ceaucescu of Romania last month cut short an official visit to Canada after he was told that he could not visit the Pickering nuclear reactor plant because of a labour dispute. He did, however, visit Hydro-Quebec's nuclear power station at Becancour, which uses Candu reactors, similar to those the official handouts somewhat misleadingly referred to the "two plants now under construction in Romania".

Romania signed contracts for the purchase of two Candu reactors in 1978 and 1981, and has options on up to eight more. But the deal has been beset by difficulties; firstly by Romania's lack of hard currency, which makes necessary a counter-trade (barter) arrangement, and then because the Romanians insisted on dealing with all the Canadian subcontractors separately, instead of negotiating with Atomic Energy of Canada Ltd (AEC). As a result, although the proposed site at Cernavoda in Romania is being prepared to receive five reactors, hardly any equipment has yet arrived from Canada. The 4,500 Romanian on-site workers are simply carrying out preliminary construction work, including the containment buildings for the first two Candus. (Containment buildings are a rarity in East European power stations; Soviet doctrine considers them a Western device to raise costs, although the Hungarians insisted on having them for their Soviet-produced reactors at Paks.)

Mr Ceaucescu is clearly anxious to get the Candus installed as soon as possible. Although Romania is an oil-producing

country, heavy investment in the petrochemical industry during the 1970s has turned it into a net energy importer. By the early 1980s there was an energy crisis, and the severe winter of 1984-85 meant prolonged power cuts for both domestic and industrial users and a ban on all private motoring. Not surprisingly, the Party daily, *Scinteia*, on the eve of the President's visit to Canada, hailed the Candu as the answer to Romania's energy problems.

The project is, of course, badly behind schedule. The first two Cernavoda Candus were meant to be operational by 1985; instead, the first major equipment will not be delivered until the autumn, and the primary cooling pumps, from Bingham Willamette of British Columbia, will not be delivered for two or three years.

But Canada too needs the Romanian contract to save jobs in its nuclear industry. Although the existing deals provide for Romania gradually to absorb Canadian technology and eventually to build its own Candu reactors (paying a royalty on the first fifteen), the Canadian content of the third and fourth Cernavoda reactors, though less than the first two, would still be significant. Furthermore, after prolonged negotiations, the two sides have signed a third-market contract, by which (subject to Canadian requirements on non-proliferation), Romania would cooperate with Canada in selling Candu reactors to other countries, particularly in the socialist and non-aligned blocs where Romanian contacts could be particularly useful.

Vera Rich

European Communities

More attention for basic research?

Brussels

Karl-Heinz Narjes, research commissioner of the European Economic Community (EEC), seems to be edging towards a more balanced programme of Community research. At an informal meeting of research ministers in Rome last week, he urged that industrial competitiveness should dominate the programme to the exclusion of research in fields with fewer technological benefits.

As things are, the EEC programme is dominated by schemes for improving European competitiveness vis-à-vis the United States and Japan (notably the ESPRIT and BRITE programmes). But Narjes stressed last week that there are "yawning gaps" in agricultural research, for example, especially the reduction of surpluses, and the improvement of food quality and of animal production. There is no research at all in fisheries and aquaculture, he said.

The interim report on the current framework programme, which will run until 1987, also talks of a serious imbalance between nuclear and non-nuclear objectives, and says that schemes for stimulating co-operation and interchange have been more "gradual" than expected.

The aim of the presentation was to start EEC research ministers thinking about their priorities for future Community research on the eve of the review of the framework programme, planned for the second half of 1985. That debate will be preceded by a comparison of national and Community scientific policies by the COPOL (Coordination of National Policies) group meeting in June 1985.

The Commission's own proposals for future areas of research include health care and technology, the management of energy resources, training, European cooperation, the mobility of researchers (inside and outside the Community) and the more effective use of scientific equipment. Moreover, the Forecasting and Assessment in Science and Technology Group, whose preliminary conclusions are due in mid-1985, wants more research of an economic character into companies and services.

Whether ministers will accept these proposals will depend not least on what they are prepared to spend on research out of the Community budget and their own resources. By March 1985, the activities of the current programme had cost 2,685 million European Currency Units (ECU, 1 ECU = £0.60) out of a total of 3,750 million ECU budgeted until 1987 — three per cent of the total Community budget. In 1984, ministers committed themselves to increase research funding. Both Jacques Delors, the president of the Commission, and Mr Narjes would like to see that figure doubled for the second framework programme covering 1988–1991/92.

The meeting in Rome also gave ministers

a foretaste of the agenda for the formal EEC research council scheduled for 4 June in Luxembourg.

● **Telecommunications research.** The European Commission is keen to get a speedy decision on its proposals for research in advanced communications in Europe (RJCE, see *Nature* 314, 209; 1985) but the United Kingdom, France and West Germany are sceptical, asking for a clear definition before embarking on research.

● **Tritium-handling laboratory.** To support the thermonuclear fusion programme, EEC has agreed to build a laboratory at the Joint Research Centre at Ispra in Italy, but Britain and France continue to doubt its siting and timeliness, given that the Next European Torus (NET), for which the tritium would be needed, will not be ready

French research

CNRS to concentrate resources

THE French national research council (CNRS) is to devote 40 per cent of its resources to twenty newly defined "strategic objectives", according to the CNRS 5–7 year plan published last week. The twenty disciplines include many fundamental subjects, such as nuclear and particle physics, but these are sometimes combined in significant new associations. Thus, the "sciences of the Universe" combine in one objective with particle and nuclear physics. Other objectives specify boundary-breaking explicitly, such as that named as "the interaction between and chemistry biology". It has long been a passion of the CNRS director Pierre Papon to forge unity among the highly divided disciplines in France; the CNRS council seems now to have given him the hammer and anvil he needs.

The new objectives also stress long-range applied science, such as Earth observation from satellites (in anticipation of the launch of the French SPOT satellite later this year), process engineering, biotechnology and so on, but Papon said last week that their selection is not a threat to the basic sciences. Many of them, he said, are in any case basic science and "it is not a major problem to defend basic science in France".

According to Papon, the real objective of the CNRS "list of 20" is "to unify science". He says he was struck, during a European tour last year, by the many links between chemistry and biology departments in Britain, whereas there is just one such department in France. Unification of disciplines is also more advanced in Sweden and Switzerland. Research in France is more fractured than elsewhere, though many modern scientific problems are multidisciplinary.

Mobility, both among disciplines and

until the 1990s, and that the site for that machine has not yet been chosen.

● **European Synchrotron Facility.** In exchange for its agreement to the proposed site at Grenoble, Denmark is seeking a marine research institute in its own country.

● **Initiative for Research in Informatics applied to Society (IRIS).** This proposal, put forward by the Italians in Rome, would encourage companies to develop new products and social services in fields such as medicine and agriculture. The proposal has been greeted with scepticism by the United Kingdom, but the Commission has agreed to a preliminary study.

Ministers also discussed proposals by the European Commission for greater control of state aid in the member states. Such control over national research programmes would be particularly unwelcome in countries such as the United Kingdom, France, West Germany and Italy.

Anna Lubinska

geographically, is poor, but Papon says CNRS is working with the ministry of research and technology to solve many purely administrative and practical obstacles, such as finding a job for a spouse. Last year there was a growth of some 15 per cent in the number of CNRS researchers moving into industry for a spell of three years or more. This might be "slow progress", but there is a "psychological problem" said Papon. Last year, one in two hundred CNRS staff made a move to industry. "It would be nice to double the figure."

Robert Walgate

Areas of research

THE research areas that CNRS will now emphasize are:

- The universe, atomic nucleus and elementary particles (some of the costs of these disciplines are borne outside CNRS).
- Mathematics.
- Optics.
- Materials sciences.
- *Filière électronique* (science applied to the microelectronics industry).
- Plasma physics and nuclear fusion.
- Energy and thermodynamics.
- Process engineering.
- New methods in earth sciences.
- Dynamics of the ocean systems.
- Remote sensing.
- Continental water.
- Understanding chemical reactions.
- Chemistry-biology interaction.
- Biotechnology.
- Neurosciences.
- Sciences of evolution.
- Sciences of communication.
- Employment, work and technology.
- Urbanism, architecture and society.
- Large scientific equipment.

UNIDO biotechnology

Twinned-centre jumps legal gun

New Delhi

DESPITE legal and other hurdles, the United Nations Industrial Development Organization (UNIDO) is trying hard to put its International Centre for Genetic Engineering and Biotechnology (ICGEB) on an operational footing this year. The sixth meeting of the ICGEB preparatory committee held last month in New Delhi has decided to launch the centre's scientific activities without waiting for the ratification of the ICGEB statutes by member countries.

So far, 36 nations have agreed to be members of the first international biotechnology centre, but only Iraq has ratified the statutes. Mr Adolfo Teylhardat of Venezuela, chairman of the preparatory committee, says the statutes must be ratified by at least 24 countries before ICGEB becomes a legal entity.

UNIDO mooted the idea of the centre in 1981 with the aim of bringing the benefits of biotechnology to bear on problems unique to the developing world. By a decision in 1984, ICGEB is to have two components, one located in New Delhi and the other in Trieste, Italy. The host countries have yet to ratify the statutes and finalize the agreements with ICGEB.

These legal hurdles are not expected to hold up the activities of the centre. The New Delhi meeting, attended by 25 countries, decided formally to begin the centre's work with a workshop on plant biotechnology in New Delhi in September. This will be followed by another workshop at Trieste on a topic yet to be decided.

In establishing ICGEB, UNIDO is being guided by an panel of eminent scientific advisers (PSA). One of the panel's suggestions, based on the experience of the Salk Institute in the United States, was to hire between five and ten distinguished scientists to serve as non-resident members of the faculty of the centre. This suggestion, though on the agenda, was not followed up at the New Delhi meeting. Nor was any decision taken on the selection of a director for the centre. Teylhardat said that ICGEB is looking for a single director for its two components who would be "a scientist of world repute with managerial experience". Selection has been deferred until the next meeting of the preparatory committee in Havana in November.

A major recommendation of PSA is that excellence and not nationality should be the basis for selection of the director and other senior staff, who need not necessarily be from member countries. Except for Greece, Italy and Spain, most of the ICGEB members are developing countries with little infrastructure or trained manpower for genetic engineering research. For this reason, most of the staff is initially expected to be from developed nations.

The ICGEB statutes provide certain im-

munity and privileges to internationally recruited staff that will not be available to the locally recruited scientists. This, together with the proposed discriminatory salary packages (\$50,000 and above for international staff) may create problems, Indian scientists fear. India has made it clear that its contributions to the centre in New Delhi would not cover the salaries of such international staff members, the cost of which will be met with funds from Italy.

Although no countries except the hosts have pledged funds, UNIDO says the operation of ICGEB for its first 30 months is assured by offers of a total \$24 million by India and Italy. ICGEB's operational cost for the first five years is estimated at \$47.8 million. It is hoped that the balance will be covered by initial and yearly contributions from members and others. Laboratory facilities are yet to be set up in Trieste or New Delhi. India has allocated \$4.3 million for 1985-86 for permanent facilities on 25 acres of land and temporary laboratory space in the Jawaharlal Nehru University.

US science budget

Deficit-cutting means freeze

Washington

RESEARCH is unlikely to emerge unscathed from Congress's efforts to reduce the federal budget deficit despite its relatively privileged position in the Reagan administration's list of priorities. The Senate is considering proposals to hold spending on various scientific and space programmes at 1985 levels. The agencies affected include the National Science Foundation (NSF), the National Aeronautics and Space Administration (NASA) and the Department of Energy. The House of Representatives, for its part, has refused to authorize the administration's proposed 4.4 per cent budget increase for NSF. Earlier this month, the House of Representatives similarly refused to authorize an increase for NASA (see *Nature* 11 April, p.488).

The proposal before the Senate to freeze spending on general scientific and space research, known as function 250, is part of the deficit-reduction package agreed last month between Senate Republican leaders and the White House, which was still being debated last week. Were the package to be adopted by the full Senate, it would become politically almost impossible for Senate committees to authorize budget increases for scientific agencies.

The House action in freezing NSF's budget would keep the authorization at \$1,501.8 million. The \$67.6 million saved would come from research (\$58.3 million) and from the US Antarctic programme (\$9.3 million). The House Science and Technology Committee had earlier recom-

Under existing arrangements, the New Delhi component of ICGEB will conduct research on biotechnology applied to agriculture, human and animal health whereas the Trieste half will focus on industrial microbiology — the conversion of biomass into fuel, bio-recovery of hydrocarbons and protein engineering. Trieste will have full facilities for developing industrially viable technologies based on genetically manufactured organisms, for scaling up technologies developed in New Delhi and a central computer facility.

By the time the centre reaches full operating capacity, each site will have a team of 30 scientists, 20 postdoctoral fellows, 30 technicians and 30 trainees from developing countries. According to Teylhardat, it will serve as an international channel for the flow of information on biotechnology from industrialized to developing nations. A gene bank containing genetic stocks and information is also planned.

Teylhardat said that ICGEB will also be the centre of a network of affiliated regional and national research institutes for biotechnology research. The New Delhi meeting considered requests for affiliation status from ten countries, including Egypt and China.

K.S. Jayaraman

mended going along with the President's proposed budget of \$1,569.4 million, but during the debate in the House, committee members offered the cut, which was accepted by a large majority.

The House bill (HR 1210) would also transfer to the director of NSF the authority to appoint assistant directors. In the past, there have been long delays in filling these posts because of the requirements that assistant directors be appointed by the President and then confirmed by the Senate. The bill also enjoins the NSF to place "emphasis" on acid rain research. The Senate has not yet voted on the NSF budget authorization, but the Commerce Committee, one of two Senate committees vying for jurisdiction over the foundation, last week recommended a budget increase broadly in line with the administration's request. The Labor and Human Resources Committee will tackle the NSF budget later this month, after which the Senate will vote; a conference committee to resolve differences with the House is likely.

Arguments in the authorization committee are partly political posturing, as it is the appropriations committees that finally decide what money goes where. The Senate has been unable to pass an NSF authorization bill for the past four years, and few have missed it; if, however, such a bill were passed by both Houses that implied a budget freeze, it would be technically difficult and politically risky for the appropriations committees to fund a budget increase.

Tim Beardsley

Australian science

Plans laid for new lobby

Canberra

AUSTRALIAN scientists and technologists have taken the first step towards the formation of their own national lobby group, to take its place alongside the others crowding the political landscape. Almost every other interest group seems to have a lobbying arm. But Australian scientists, although producing two per cent of the world's published scientific literature from a population base comprising less than one-half of one per cent of the total, have only now realized that a performance-enhanced mousetrap, say, will not automatically be taken up by industry, or even recognized by the media, let alone by politicians. This is why, with a few notable exceptions, scientists are a fugitive breed in Australia, misunderstood when not actually invisible. Last week's nucleation at a meeting in the Australian Academy of Science of a body to lobby government and spread the good word generally may help to rectify the situation.

The spur has been the August 1984 federal budget, in which publicly supported science was cut back by five per cent. With recruiting among academics and in the Commonwealth Scientific and Industrial Research Organization (CSIRO) already at a standstill, it began to look as if even scientists holding permanent jobs would have their research hindered. An *ad hoc* group of eight scientific societies expressed disquiet about the cuts. The Academy of Science established a 16-member National Committee for the Promotion of Science and Technology under the chairmanship of Professor Max Bennett, a neurobiologist from the University of Sydney.

The immediate outcome was the National Meeting of Concern on Science and Technology at which the elected leaders of 68 scientific and industrial societies and two academies representing about 80,000 members voted to elect an interim federation committee. That committee's first task is to explore the setting up of two secretariats, one of which will be a small professional lobbying group based in Canberra to facilitate communication between government and the scientific community, especially at budget time.

The other committee (probably based in Sydney), will provide a science and technology information service modelled on the Media Resource Service of the US Scientists Institute for Public Information. Through it, people in the news media, education and politics throughout Australia would be put in touch with experts for comment on scientific matters. The twelve-member interim committee, under the chairmanship of Professor Fred Smith, a physicist at Monash University, is due to report in six months.

A mood of optimism pervaded the meeting, but there were objections that

political consensus may be difficult to achieve between academic, CSIRO and industrial scientists, particularly because the specialist societies (some with their own lobbying apparatus) differ greatly in size.

The federal Minister for Industry, Technology and Commerce, Senator John Button, said that, although he welcomed the possibility of discussions with such widely representative groups, not all basic research could be convincingly sold as seed corn for the future. In the interests of industrial restructuring, he said, public financing "may not be just where you are used to it being".

The Minister for Science, Mr Barry Jones, was absent from Australia during the meeting, but in February this year

something was salvaged from the 1984 budget wreck in the form of one of his favourite projects — the Commission for the Future.

The aim of the new body is also to raise public awareness and debate about science and technology, especially about the ways in which technological change will affect the lives of ordinary people, and the choices they must make in education and employment. The commission is headed by Mr Philip Adams, an advertising executive and all-round pundit, who is an old friend of Mr Jones.

The scene is set, then, for the emergence of science (and some scientists) from the shadows where they have languished politically for so long. Noting how badly the job of advocacy has been done in the past, Professor Smith says that, realistically, a major impact may not be felt for five years.

Jeffrey Sellar

Famine

Remedy not by bread alone

Anaheim, California

THE more equitable distribution of available food supplies is not, by itself, the best way of avoiding famine. This is what a gathering of biologists was repeatedly told last week. And so it is a matter of urgency that biologists should pay more attention to the prediction, the causes of onset and the prevention of famine.

Organized by Dr Barry Bloom of the Albert Einstein College of Medicine, the symposium was belatedly and exceptionally tagged onto the 300 or so sessions already jamming the programme of this year's meeting in Anaheim of the Federation of American Societies for Experimental Biology (FASEB). Bloom, an immunologist with a special interest in leprosy, was prompted to organize the symposium by the present African famines. "We have learned so few lessons", he says, "that each new catastrophe tends to repeat some of the classic errors from the past."

One priority in research should be to develop criteria for predicting famines in time to prevent them. "In most circumstances, one is likely to have one group of people politically or emotionally interested in asserting that the situation is worse than it is and another arguing the opposite", says Bloom.

That theme was taken up by John Mellor of the International Food Policy Research Institute in Washington, DC, who argued that if more famine-prone countries had democratic governments and a free press, we might learn of problems in time to prevent them. Foreign aid must be channelled into transportation systems so that food can be distributed, he said, and into the training of an elite of agricultural personnel. The priority for scientific research should be the mounting of a green revolution in Africa.

Although confident that scientists could

produce the means for such a revolution, Dr Robert Goldman, of the plant biotechnology company Calgene, wondered if it would be put into effect. The science that fuelled the green revolution that saved India and Mexico from famine was developed in the public sector and applied without profit. Now that the impetus for plant improvement has passed to the private sector, particularly agricultural biotechnology companies, the developing countries may be less able to win benefits.

Famine is, in any case, the end result of a chronic condition of malnutrition. Eight hundred million people worldwide are calory-deficient, and "it seems likely that whole societies have been forced to adapt to this", says Dr Neville Scrimshaw of the Massachusetts Institute of Technology. Adaptation means reducing physical activity, which can have severe social consequences. Less obvious deficiencies of iron, vitamin A and iodine are widespread, debilitating and make people more prone to infections which, in turn, can exacerbate malnutrition.

The downward spiral that ends in famine is best broken by immunology, in the opinion of Stephen Joseph of UNICEF. Calling for more research into both vaccine development and the inter-relationship between immunity, nutrition, infection and growth, he said that scientific research must go hand in hand with improved management of preventive programmes and better education of those who should benefit.

After the symposium, Dr Bloom complained that no more than 200 people had attended. Up to 15,000 people were expected to register for FASEB, and although the famine symposium was held the evening before the full meeting got under way, 2,000 people turned up next day at a lecture on the molecules involved in lymphocyte activation.

Peter Newmark

AIDS

Undesirable import to Japan

Tokyo

WITH Japan under pressure to increase imports of US goods, it is ironical that one product Japan imports freely from the United States may have brought acquired immune deficiency syndrome (AIDS). According to last month's *Taisha*, the Japanese medical journal, two now-dead Japanese haemophiliacs were killed by AIDS, probably by infected blood plasma, almost certainly imported from the United States.

One of the victims, aged 48, died in July 1983 and the other in November last year, aged 62. The victims, the first from AIDS in Japan, showed an extraordinary loss of immunity, and the presence of antibodies to the virus HTLV-III has been confirmed.

No cases of AIDS have yet been reported among Japan's homosexual community, but a confirmed Japanese homosexual now living in the United States who returned to Japan in mid-January was diagnosed as an AIDS sufferer. The AIDS Research and Study Committee of the Health and Welfare Ministry says there is little chance that the man infected others in Japan.

The emergence of AIDS cases among Japanese haemophiliacs seems to be attributable to Japan's almost insatiable demand for US blood plasma. Despite one of the highest rates of blood donation in the world, Japan depends on the United States for nearly 90 per cent of the plasma it consumes, amounting to 300,000 litres a year. Demand for plasma-protein fractions and blood-component preparations is rising, for the latter by 15 per cent a year.

So why is Japan so dependent on blood plasma imports when there were more than eight million blood donations last year? Officials at the Health and Welfare Ministry explain that donors in Japan give only 200 ml of blood on each occasion, compared with about 400 ml in other countries. But a Japan Red Cross official says that in the United States in 1982, just under six million donations of about 450 ml were made, suggesting that *per capita* blood donation in the two countries is similar.

Dependency on blood imports is in part a legacy of the system before 1982. Blood donations were then saved almost exclusively for the use of the donor himself, who would resist the idea of volunteering blood for others, probably for the reasons that also make Japanese reluctant to donate organs for transplant (see *Nature* 313, 338; 1985).

The Health and Welfare Ministry's planned solution of this vexing problem is threefold. First, Japanese donors will be asked to donate 400 ml instead of the present 200 ml. Second, plasmapheresis, which separates the plasma and returns red cells to the donor, will be more widely used. Third, doctors will be asked to avoid un-

necessary wastage of plasma. In the meantime, however, imports will continue and there are no plans to screen imported blood for AIDS.

All of this must provide small comfort for Japan's haemophiliacs. The article in *Taisha* even reports that the blood of 23 out of a random sample of 50 Japanese haemophiliacs contained antibodies to the virus HTLV-III, which suggests that a large proportion of Japan's 5,000 haemophiliacs may have already been exposed.

Even if imports stopped, there would be

another danger. Drs Kazuo Ohkouchi and Hiroyuki Sato of Kyushu University in southern Japan have now shown that the virus HTLV-I, which causes adult T-cell leukaemia (ATL), can be transmitted by blood transfusion. ATL is endemic in southern Japan, where approximately 20,000 carriers are found among blood donors each year. Dr R.C. Gallo, chief of the tumour cell laboratory at the US National Cancer Institute, has even suggested that an AIDS-like epidemic of ATL may be imminent. Clearly, the management of blood supplies in Japan is a matter that in future will come under increasing scrutiny by the eyes of the world.

David Swinbanks

ESA, Japan agree on participation in planned US space station

EUROPE formally agreed last week to join the United States in the development of a space station, but details such as who spends how much on what remain to be decided. The European Space Agency (ESA) will now sign a memorandum of understanding with the National Aeronautics and Space Administration (NASA) which requires that the design of the space station, with costs and commitments, must be ready by the end of 1986. This phase of the work will cost ESA \$65 million.

NASA has already signed a similar agreement with Canada, which aims to provide robot-servicing equipment at a cost of some hundreds of millions of dollars, and is due to sign another with Japan this month (see above).

The European contribution may finally be some \$2,000 million, which should buy a pressurized module that could be used as a manned laboratory (called Columbus), a number of free-flying payload carriers for both low-inclination and polar orbits, a servicing vehicle and a resources module that could provide both the pressurized module and the payload carriers with electric power, cooling and stabilizing systems.

These housekeeping elements have been included in ESA's plan at French insistence, and would enable the European modules to fly separately from the US space station if necessary, although (for the moment) Europe would still have to rely on the US space shuttle to put its astronauts into orbit. France, however, has plans for a mini-shuttle (Hermes) and in Britain, British Aerospace has described the concept of a horizontal take-off and landing spacecraft (HOTOL). Either of these could ultimately provide people with independent access to space.

There are more immediate issues still to be resolved. US fears that technologies transferred to Europe would leak to the Soviet Union engender corresponding European fears about access to US science, technology and data. There are also the issues of who pays the still unknown

maintenance and running costs of the station, the legal framework for deciding patent rights arising from work on the station and the management and decision-making structure of the enterprise.

Robert Walgate

Tokyo

JAPAN has decided to take part in the US project to build a manned space station. James Beggs, head of the US National Aeronautics Space Administration (NASA), who has made earlier trips to Japan to promote the space station, will return to Tokyo for the signing on 9 May of an agreement between NASA and Japan's Science and Technology Agency (STA).

The initial agreement will cover a two-year period of design work and will provide Japan with technical information on the space station. STA is refusing to release details of the budget until after the signing, but it is known that some 1,400 million yen (\$5.6 million) has been earmarked for the project in this year's budget.

Informally, a consensus seems to have been reached that Japan should take the opportunity to master new technology and build one whole experimental module of the space station. A subcommittee of STA's Space Development Committee has recommended that it consist of three parts: a pressurized cabin ten metres long and four metres in diameter for use in experiments on new materials and new pharmaceuticals more easily produced in gravity-free conditions, a service platform for testing of telecommunications equipment in outer space and a storage cabin. Total design and construction expenses are likely to exceed Y200-300 thousand million million (\$800 million).

Industry has already expressed considerable interest in the space station and several large corporations, including Mitsubishi, Mitsui, Sumitomo and Fuyo, are now merging their respective study groups to find the best way to win commercial advantage from the station. Alun Anderson

Popularization

Royal Society asks for more

THE Royal Society, the premier British scientific society, is to be given a hefty nudge towards the popularization of science, to judge by a meeting on the "public understanding of science" held under the auspices of the British Association for the Advancement of Science in London on Monday.

The meeting was convened by Sir Hans Kornberg, British Association president, to provide a debating forum for the Royal Society *ad hoc* committee on public understanding. And while many old chestnuts were brought out on Monday and re-roasted, Dr Walter Bodmer, chairman of the Royal Society committee, used the occasion to demonstrate his committee's determination that its work would not gather dust on any shelf.

The need is certainly great. In Britain, on the whole, there is little sign of a scientific lobby, nor of any determined effort by the universities or research councils to educate or even entertain the public in science. Bodmer and his committee wish this situation to change, and by starting at the Royal Society, traditionally a somewhat austere institution whose motto is "*nullus in verba*" (nothing in words) they are certainly taking the bull by the horns. Dr Bodmer hoped the Royal could provide a lead which others could follow.

During Monday's meeting, Dr Bodmer, director of the Imperial Cancer Research Fund, received support from many quarters both in the media and in government. Sir Robin Nicholson, chief scientist in the Cabinet Office, claimed it is now almost impossible to find a political issue without scientific or technical content, and that "90 per cent of his time" is used in advising government on these matters (and only 10 per cent on science itself). Science and technology are now central to many foreign policy issues, said Sir Robin, in the way that international trade took centre stage 25 years ago. But "the attention span grows shorter as one gets older" so now cabinet ministers can only cope with five-minute briefings or three pages of A4. The same is true of leaders of industry. But scientists, for good reasons, are trained to write *in extenso* and giving equal weight to many ideas. This is completely at odds with what is needed in government, or in the press.

Dr Bodmer was keen to affirm that he would like the Royal to play an active role in getting scientists to approach the media. A member of the science advisory panel to the British Broadcasting Corporation (which, incidentally, reckons it shows four times as many science hours in proportion to output as does US television), Bodmer is also taking seriously advice to expand the

Royal Society's programme of science briefings, to institute rapid press conferences on scientific issues behind and in the news, to run deeper, background seminars on scientific matters, and to produce detailed condensed briefing documents for journalists.

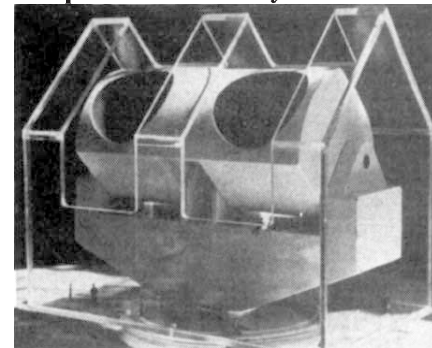
On a warning from Sir James Gowans, chairman of the Medical Research Council, that such exercises cost money and a lot of time, Dr Bodmer said he hoped the Royal Society would be able to find backing "if not from government" from industry — which should benefit in the long run from better public knowledge. His committee's report is due in a few months' time.

Robert Walgate

Big telescope race

Washington

OHIO State University and the University of Arizona have decided to raise a total of \$25–30 million for an 8-metre monolithic mirror instrument to be built on Mount Graham in southeastern Arizona. If others can be persuaded to join the project, the telescope may even be equipped with two 8-metre mirrors, which could be combined for optical interferometry.



The universities have not yet identified the private sponsors who they hope will supply most of the capital cost of the instrument. Both universities have, however, recently launched large fund-raising efforts, with a target of \$100 million at Arizona and \$250 million at Ohio. Astronomers at both universities are confident that the money will be found.

The telescope, which would also operate at infrared wavelengths, will have computer links allowing it to be operated remotely from both Arizona and Ohio. The mirror or mirrors will be made using a technique developed by Roger Angel at Arizona (see *Nature* 295, 651; 1982) in which mirrors of honeycomb structure are spun-cast into a parabolic shape. A new computer-actuated lap allows very deep dishes to be polished with a focal ratio of $f/1$, so that telescopes can be built that are much smaller and cheaper than hitherto. A 1.8-metre mirror made recently using the new technique "came out beautifully", according to Dr Angel. Preliminary design work on the new telescope is due to start later this year; construction might take a further four years.

Tim Beardsley

Animal experimentation

Alternatives neglected

Washington

A STUDY of animal experimentation by the National Academy of Sciences* suggests that institutional barriers within the biomedical community in general and the National Institutes of Health (NIH) in particular may discourage researchers from developing models of human disease and physiology based on non-mammalian organisms. The study, which was commissioned by NIH, does not directly criticize NIH's funding patterns; it does, however, recommend that the agency should support good research "without taxonomic or phylogenetic bias" and consider directing funds specifically to the development of promising model systems. Harold Morowitz of Yale University, chairman of the study, said there is "anecdotal evidence that it is difficult for people working on unconventional models to get funding". The natural bias has been in favour of using organisms closest to man, he said; unconventional proposals submitted to NIH for support do not usually fit into the mandate of any of the standard study sections, and are reviewed by *ad hoc* committees that are not always sympathetic.

Morowitz said that researchers working

in unconventional systems often report that they managed to get started only by "bootlegging" funds from other more conventional research projects.

The study notes that as biological understanding shifts more and more to the molecular level, increasingly the best model for human disease and physiology is not a closely related whole organism that displays a similar macroscopic feature; rather it is often many organisms, each an analogue of a fundamental biological principle involved. One example cited in the study is research on the human disease myasthenia gravis, a neuromuscular disorder. Six different animal models have been used to discover its cause: an autoimmune attack on acetylcholine receptors at neuromuscular junctions. Studies on chick muscle cells in culture showed that a snake venom binds specifically to these receptors; the venom was then used to demonstrate that patients had a striking shortage of receptors. A cobra neurotoxin that blocks the receptors was administered to rats to see if a shortage of receptors could explain the observed disease symptoms. And to see if an autoimmune response could be involved, rabbits were immunized with receptor cells isolated from invertebrate species.

Stephen Budiansky

*Models for Biomedical Research, National Academy Press, Washington, DC, 1985 \$18.50.

Changing the objectives

SIR — Your leading article of 7 March (p.1) deals with policies for UK science, including astronomy, in the context of fixed or declining resources. You say: "... one, or the other, or perhaps both of the observatories must go ...". You suggest that the "good work these laboratories do ... could be transferred intact to universities". Another problem is that experimental facilities have become "institutionalized".

I suggest that a contract-based laboratory might be set up in the general subject area of physical technology, of which astronomical technology would be part. It would absorb some staff and property of the Royal Greenwich Observatory and the Royal Observatory, Edinburgh. It should probably not absorb the Nautical Almanac Office and historical collections of the observatories, which need limited resources and could continue independently.

This suggestion is intended to be more far-reaching than simply merging the observatories. The organization is not going to be a literal observatory because the relevant instruments are overseas. It is more of a research and development organization, a design office and a workshop. Although such a new laboratory would start by working on instruments for astronomy, that need be only one of its jobs. There seems little to be gained from total specialization. The aim would be to add versatility and university contacts and to give the establishment responsibility for its own fate. Versatility in the face of change is relevant to present circumstances. There could be many other gains. Developments in one area of technology such as astronomical instrumentation need inputs from other subjects and in turn ought to be generally disseminated — that process would also be helped. A vigorous working environment is foreseeable.

The question must be asked: would such a laboratory take money or talent needed elsewhere? The answer must lie in the way it is run. Its income would not be unfair if it was gained by offering a service that could be refused. Such operations would be based on contracts.

For a new laboratory, the foundation itself could take the form of a contract or agreement between the Department of Education and Science and the management authority, which could be a university. A notable example is the vast Jet Propulsion Laboratory in Pasadena, which is managed by the California Institute of Technology on a contract. To maintain the service principle, the foundation agreement itself should not provide much of the running expenses for technology. The foundation could reasonably support some scientists. The research councils should probably not be legally involved in the foundation as the laboratory would be one of their contractors.

It should be appreciated that the observatories' research astronomers are few in number and their continued support is not a problem that can relevantly be compared with that of providing and operating the larger research facilities which are basically for universities. The practice of astronomy has been the attractive feature of the observatories for many staff who have eventually worked in technology. The total contribution of the observatories' astronomers to the provision of facilities, as well as to the provision of fundamental data and to research programmes, is such that they can surely be included in the framework proposed. Some posts would be part of the foundation agreement, some would be Science and Engineering Research Council (SERC) fellowships applied for in the usual way, some posts including direct responsibility for instrumentation would be supported by contracts. It would be open to the organization to use any net income to employ extra scientists.

Such a laboratory would be clearly distinguishable from others. The differences would be in the combination of the university connection, contract operation, flexibility in the type of work, the laboratory and the university together seeking work

and the origins in astronomical technology (mainly optics, detectors, computing and mechanical engineering).

The operation of existing research facilities and building of new ones would be on the basis of a number of separate long or short-term contracts. Scientists requiring such work would probably act through SERC, using either a grant or a direct contract from the council for common facilities. Ultimately, these contracts could be placed or not as required and could be refused by the laboratory: cancellation would be regulated by the contract.

The laboratory would therefore have the need and the opportunity to support itself (as distinct from subsidizing the facilities). This it would do by seeking government and commercial work which should not be circumscribed. The laboratory could sell its services, software and the rights to its inventions.

From the scientists' point of view, much would depend on the continuing goodwill of the managing university. Success would depend on the local management and staff, and sanctions would be available through the contracts.

RICHARD G. BINGHAM

*Brackenholme,
Jenners Lane,
Wartling, Hailsham,
East Sussex BN27 1RU, UK*

Plight of refusniks

SIR — Ten years ago, on 12 March 1975, my family applied for exit visas from the Soviet Union for the first time. I was immediately removed from my position in the mathematical department of the Institute for Chemical Physics of the Academy of Sciences of the USSR. At the same time, a total boycott was organized against my family in the scientific centre of the academy in Chernogolovka.

For ten years I have been prohibited from attending seminars, conferences and symposia. Correspondence from abroad is not delivered to me. I do not receive scientific journals, including those of the American Physical Society and the American Mathematical Society. The Academy of Sciences presses my foreign colleagues to abandon attempts to meet me and then to take part in the scientific boycott against me.

For ten years we have been stubbornly making efforts to get exit visas. In 1979, representatives of the academy, academician N. Emanuel, Academician N. Semenov and others, officially stated that the academy had no objections to my departure, and we were informed that the Soviet leaders had made an official decision to give our family exit visas. But this decision has not been implemented.

My numerous meetings with the general secretary of the academy, Academician G. Scrjabin, showed that the academy blocked the fulfilment of this decision.

The situation has abruptly become worse

after negotiations between the Academy of Sciences of the USSR and the National Academy of Sciences of the United States. Academician Scrjabin and Academician Yu. Ovchinnikov, a vice president of the academy, immediately refused to discuss my problem with me. And Academician E. Velikhov, another vice president of the academy, refused even to speak with me.

SOLOMON AL'BER

Moscow, Soviet Union.

Abstract policy

SIR — In regard to the previous correspondence on the abstract policy of the Biosciences Information Service (BIOSIS) (H. E. Kennedy, *Nature* 310, 96; 1984), I would agree with those favouring their inclusion. I would point out, however, that the only thing worse than no abstract is a reference to one when there is no obvious way of obtaining a copy. BIOSIS apparently does not intend to offer OATS (original article tear-sheet service) as does the Chemical Abstracts Service.

Would it be too much to ask that some record be provided as to which libraries hold these abstract volumes in a manner similar to CASSI (Chemical Abstracts Service Source Index) or that they be deposited in the British Library Lending Division or a library which reports its holdings to the Online Computer Library Center?

DANA L. ROTH

*Science and Engineering Libraries,
California Institute of Technology,
Pasadena, California 91125, USA*

Industrial collaboration on show

Professional people do not spend time as salesmen at trade fairs each year. But there is much to be learned from what went on at Hannover last week.

Hannover

ALMOST as a matter of principle of general applicability, most people agree that it is wrong to throw out the baby with the bathwater, an image that no doubt derives from the time when baths were movable and plumbing for carrying away waste water had not been developed. The difficulty in most interesting cases is to know which is which. That may be one of the reasons why British universities have for so long felt so threatened by the British government's mixture of exhortation and incentive intended to make them contribute more directly to national prosperity. More applied research means less fundamental research, as in any zero-sum game. The huge annual trade fair that ended here last week provided ample illustrations that this calculation is far too simple.

One of the most unexpected features of this trade fair is that universities take their place alongside industrial companies as exhibitors, sometimes under their own steam, sometimes under the wing of the governments that support them, the *Länder* in West Germany. This year, one of the most important shows of this kind was that provided by the universities of North-Rhine Westphalia, the regional government whose policy on higher education was the most expansionist of all when the whole of higher education in West Germany was growing rapidly fifteen or twenty years ago. Lower Saxony, the local *Land*, was also represented, as was Berlin, universities such as Heidelberg as well as a host of publicly supported laboratories whose work is, in part at least, academic in character. At a guess, there would have been more than 200 people with academic positions or pretensions manning stands at all times during the eight days of the fair.

Why were they there? There is no secret. As with the strictly commercial exhibitors, money is the objective, either an agreement with an industrial company for the commercial development of a device or such a powerful expression of industrial interest that further funds for a research project can be had from one of the public sources of research support, the *Deutsche Forschungsgemeinschaft* (an all-embracing research council), the ministry of research and technology (BMFT) or, more often than one would guess, some other ministry.

The ingenuity and diversity of what these academic people have had to show has been compelling, which is attested by the way in which the university exhibits have gener-

ally been occupied with visitors engaged on deep discussions. (Quite often there is also a certain amount of socializing, as when people from industry try to discover what their ex-teachers are doing now.) Machines and devices predominate, computer programs and applications come next, but there is also a cement of biological research, and of electronics applied to medicine, to give the show coherence.

Diversity is also represented by the variety of institutions with things to show. Some of the great universities are there, but so too are the technical colleges of West Germany (*Fachhochschulen*) and the more recent and controversial creations, the comprehensive universities of the 1960s (*Gesamthochschulen*). In interest and ingenuity, there has been little to distinguish between them. One feature of this unusual academic show is that the chief salesman (and inventor) has been a full-throated enthusiast for what he has to show who can be diverted only with difficulty into talking about more general questions, how his university is faring, or how he balances research and/or development with teaching. (The personal pronoun "he" seems generally applicable.)

The question where the idea came from in the first place seems to yield a great variety of answers. One academic inventor says he had returned from a previous Hannover fair astonished that the great engineering companies had not thought of a way of solving a simple technical problem. Many say their work has its roots in the immediate needs of a local hospital or industrial company.

Showing off the product is not exhibitionism, but an attempt to put the development concerned on a sounder footing. The academic exhibitors also agree that success will bring them kudos back home, perhaps inversely to the standing of their institution. Although the exhibitors are necessarily a self-selected group, they make it seem entirely natural that showing a device should be a part of normal academic life.

Like the academically inclined researchers from the West German national laboratories, they no doubt sense pressure from above to be doing something useful. Indeed, for the national laboratories, technology transfer has been this year's theme for the fair. Inevitably, many of their projects are on the grander scale. The surprise is that many of them are just as particular. Techniques for bonding metal and ceramic may as well be used in useful

domestic gadgets as, for example, in thermonuclear machines. But for the people from the national laboratories, where government edict may matter more directly, there is less enthusiasm for the incidental applications than for the real purpose.

The texture of the exhibition of research is the biggest surprise. It is natural that inventors should be enthusiasts with, occasionally, an air of zealotry. Predictably, the things on show range from particular devices to generally applicable techniques. The theme that holds them all together is the belief that there may be profit in what they have to show.

Everybody acknowledges that this is just a chance, but they are keen to be on show as if that were sufficient in itself; at least they will be able to compare their work with that of others, or learn something useful from their visitors.

The morals in this experience for academic communities other than West Germany are obvious and widely understood. Indeed, in places such as Britain, everybody pays lip-service to them. First, there is no necessary contradiction between membership of an academic faculty and a research interest which is better called development.

Second, there is no way in which the professional people manning these stands can be distinguished from those encountered at quite different gatherings, conferences for example; perhaps this is something distinctive about West Germany, but the machine builders seemed refreshingly interested in what, say, the particle accelerator people are up to. (To university administrators, in any case, the only achievement that counts is success.) Third, even in publicly supported systems of higher education (and the West German system institutionalizes government support for universities to a degree no longer allowed in Britain), extra funds without strings attached are translated into ordinary currency at a higher rate.

That industrial collaboration can be rewarding and respectable is understood elsewhere in Western Europe and North America. (Japan is different; universities deal privately with industry at the outset and exhibit only their successes at even grander fairs.) But while understood, there is a sense in which it is not believed. Academics who dispute that will be more convincing when they put themselves on show, at Hannover or elsewhere.

John Maddox

Population biology

Evolution of pesticide resistance

from Robert M. May

ALL natural populations contain genetic variability. The selective forces resulting from continued application of pesticides (or herbicides or antibiotics) will therefore lead to the target population eventually becoming resistant. There are now several examples of evolution at work in this way that have been studied in at least as much detail as industrial melanism in the peppered moth, the standard example of microevolution in elementary biology texts. Over 400 species of agricultural pests are now resistant to one or more pesticides, and this erosion of the early successes of chemical pesticides is reflected in the fact that in the 1940s around 7 per cent of all US agricultural crops were lost to insects, whereas the corresponding figure is now 13 per cent. Although the ultimate appearance of population-level resistance is usually inevitable, a variety of measures may be adopted to retard its appearance in particular cases.

For an enormous range of different pesticides and target organisms, the number of generations taken for a significant degree of resistance to appear exhibits a surprisingly small range of variation, typically being around 5–50 generations, despite great variation in the underlying genetic mechanisms and in the intensity and frequency of pesticide application. Conventional population genetics can shed light on this observation. Suppose, to begin with, that resistance involves only one locus with two alleles: *SS* genotypes are susceptible to the pesticide, *RR* genotypes are resistant, and *RS* heterozygotes have some intermediate degree of resistance (that is, $w_{RR} > w_{RS} > w_{SS}$, where w_{ij} is the fitness of the genotype *ij* in the presence of the pesticide; if *R* is completely recessive, $w_{RS} = w_{SS}$; if *R* is completely dominant, $w_{RS} = w_{RR}$). The resistance allele *R* will usually be at very low frequency (p_0) in the pristine population, and under pesticide application its frequency will at first tend to increase in the ratio w_{RS}/w_{SS} from generation to generation¹. The number of generations, *n*, that elapse before the *R* gene attains a frequency sufficiently large to produce noticeable resistance in the population as a whole (p_1) may thus be estimated by compounding this fitness ratio until $(w_{RS}/w_{SS})^n \sim (p_1/p_0)$.

From this very crude approximation, we can see that the number of generations required for the population to evolve resistance depends only logarithmically on fitness ratios and on p_0 . In other words, the characteristic time to evolve resistance indeed depends on such quantities as p_0 (estimates of which range from 10^{-2} to 10^{-13} in various circumstances according to R. T. Roush of Mississippi State University), the degree of dominance of the *R*

allele (usually intermediate between the extremes of recessive and dominant), the strength of selection (which ranges over several orders of magnitude, depending on the management of the particular pest/pesticide system), and other factors. When logarithms are taken, however, these enormous ranges collapse to a factor of around ten separating the further extremes, making the roughly 5–50-generation 'rule' a good deal less surprising.

Discussion at a recent symposium* revolved around the interesting details and exceptions that make the preceding crude generalizations inapplicable to specific cases. For example, is an insect pest that has evolved resistance to one pesticide likely to be faster in evolving resistance to another pesticide? May mixtures of pesticides work better than the same pesticides used in sequence? Better understanding of the molecular biology underlying the genetics of resistance can help here. Faster resistance to new pesticides is clearly likely if they are chemically similar to old ones. Less obviously, as emphasized by M. Whitten (CSIRO, Canberra) and others, the first pesticide could promote selection for mutator genes, thus speeding the evolution of subsequent resistance even to chemically dissimilar pesticides.

Although much of the discussion centred around the molecular biology of resistance, more attention to the population biology of resistance is also needed. G. P. Georgioudis (University of California, Riverside), Roush and others emphasized that migration from untreated to treated regions by susceptible individuals represents flow of susceptible genes that works against selection for resistance. This invokes echoes of the grander theme of whether geographical isolation is necessary for speciation^{2,3}; that is, under what circumstances will gene flow (migration) wash out the selective forces (pesticide application) that are tending to adapt an organism to its local environment (evolving resistance)? H. N. Comins and others have shown that the answers depend both on the magnitude of migration in relation to the strength of selection, and on the type of density dependence exhibited by the pest population^{4,7}.

In deliberately oversimplified terms, a population is said to exhibit 'perfect' density dependence if, following any disturbance, the population tends to return to some constant value in the next generation. In contrast, if an 'overcompensating' population is perturbed below its long-term

average value in one generation (for example, by pesticide application), it will tend to bounce back above this value in the next generation. A population with 'undercompensating' density dependence will tend to recover steadily and monotonically following disturbance; if depressed to low values by pesticide application in one generation, such populations will tend still to be lower than usual in the next generation. When these considerations of population dynamics are added to selection and gene flow, it is concluded that — other things being equal — pest populations with overcompensating density dependence will tend to evolve pesticide resistance faster than ones with 'perfect' density dependence, because the migration of susceptible genes into the overcompensated population that follows pesticide application has less effect than would be the case under 'perfect' density dependence⁴. Conversely, if the pest population exhibits undercompensating density dependence, the resistance-slowing effects of migration from untreated regions will be relatively more significant⁴.

In practice, the weather and other environmental factors often superimpose density-independent fluctuations that mask the density-dependent effects, at least in the short run; for this and other reasons, analyses that characterize the degree of overcompensation or undercompensation of particular insect populations are more easily carried out for laboratory than for field populations. However, whereas most field populations for which good data are available appear to exhibit undercompensating density dependence (among the exceptions, the most pronounced is the Colorado potato beetle, which is notorious for the speed with which it has developed resistance to a wide range of pesticides), the majority of laboratory populations show marked overcompensation^{8–10}. The difference, which probably derives from the many natural mortality factors that are seldom present in the laboratory, underlines the need for caution in extrapolating laboratory studies of the evolution of resistance into a field setting.

The propensity for pest species to evolve resistance more quickly than do their natural enemies has often been remarked, and creates obvious problems. Building on earlier work by Croft and others^{11,12}, B. E. Tabashnik (University of Hawaii) surveyed two possible explanations. The first is that the co-evolution between plants and phytophagous insects has preadapted the insects to the rapid evolution of detoxifying mechanisms, whereas their natural enemies have had less opportunity to preadapt. But laboratory studies show that there are no simple general patterns of this kind and that under controlled conditions the rate of evolution of resistance in prey and predator populations depends on the detailed molecular mechanisms underlying detoxification^{13,14}. This, in turn, has prompted a search for pesticides that may be less lethal for natural enemies than for

*National Academy of Sciences/National Research Council International Symposium on Pesticide Resistance Management, 27–29 November 1984, Washington D.C.

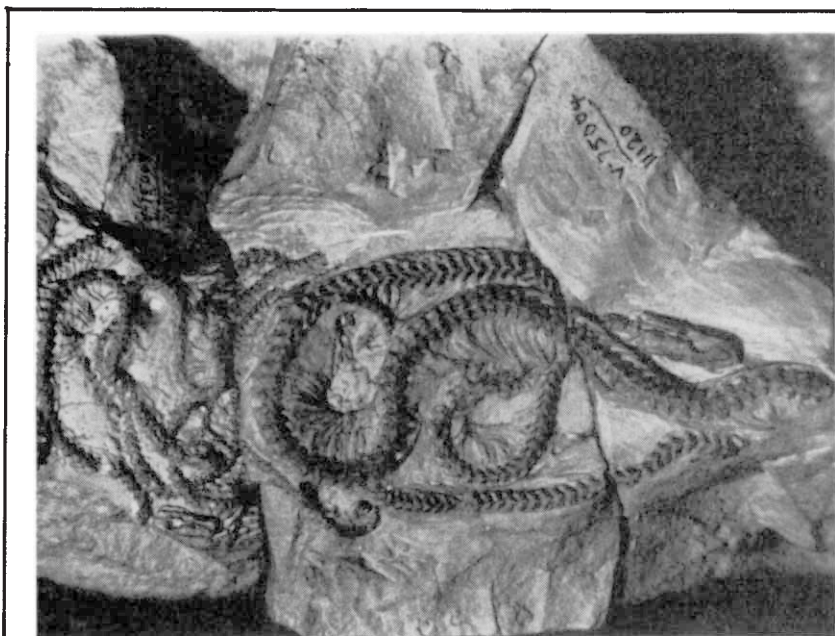
pests¹⁵⁻¹⁷, or even for the release of natural enemies that have been artificially selected for resistance to specific pesticides¹⁸.

The second possible explanation lies in the dynamics of prey/predator associations. The essentials of Tabashnik's detailed numerical simulations¹¹ can be appreciated from the qualitative arguments given above. Suppose application of a pesticide kills a large fraction of both prey and predators in the treated region: for the surviving prey, life is now relatively good (relatively free from predators), and the population is likely to increase rapidly — exhibiting overcompensating density dependence; for the surviving predators, life is relatively bad (food is harder to find), and their population will tend to recover slowly — exhibiting undercompensating density dependence. But we have just seen how the interplay between selection and gene flow is such that overcompensation tends to speed the evolution of resistance, and undercompensation to retard it. Thus, on dynamical grounds alone, we may expect that pests evolve resistance faster than do their natural enemies.

Although many well-understood resistance systems involve one locus with two alleles, with heterozygotes of intermediate fitness, other systems are possible and can pose different kinds of management questions. For instance, *RS* heterozygotes may be the most fit in the presence of the pesticide. Resistance of the Norway rat to warfarin is the clearest example, and anyone who is tired of presenting sickle-cell anaemia to classes as the canonical example of heterozygous advantage may care to note the fitness ratios of the Norway rat in Britain in the absence of warfarin are $w_{RR} : w_{RS} : w_{SS} = 0.46 : 0.77 : 1.00$, whereas in the presence of the poison they are $0.37 : 1.00 : 0.68$ (J. Greaves, MAFF, South Tolworth, Surrey). But there may also be examples of heterozygous superiority among insects, as in the resistance of the spotted root-maggot *Euxesta notata* to DDT or dieldrin¹⁹. Heterozygous superiority of this kind undercuts the simple analyses presented above²⁰.

Resistance systems that are sufficiently polygenic to warrant treatment by quantitative genetics, rather than being inherited in discrete, diallelic fashion, present interesting possibilities (S. Via, University of Iowa). For example, resistance to two different pesticides can be negatively correlated, enabling the evolution of resistance to be retarded by alternating the pesticides. Conversely, if the correlation were to be positive it would result in swift appearance of resistance to the second pesticide, following sustained use of the first. M. Uyenoyama (Duke University) discussed possible complications arising from modifier genes in two-locus systems of resistance. One possibility is the removal of fitness disadvantages in resistant genotypes in the absence of pesticide usage, if ever the resistance gene attained high frequency; if this happened, one could not

Most ancient aggregation of serpents



THE oldest record of serpent aggregation is preserved in this rare fossil from the Oligocene White River Formation of Converse County, Wyoming. The assemblage consists of three virtually complete 32 million-year-old snakes passively coiled in life-like positions. As described by Brent H. Breithaupt and David Duvall of the University of Wyoming, USA (in a paper in the press in *Rates of Behavioral Evolution*, ed. Bocout, A.J., Plenum, New York; and in a manuscript under review) the largest snake (435 mm long) belongs to the genus *Ogmophis*, whereas the smaller pair belong to the genus *Calamagras*. Both genera are common middle Tertiary forms of the subfamily Erycinae, family Boidea, and are distantly related to the North American rubber boa (*Charina*) and rosey boa (*Lichanura*). This specimen represents the most complete material known for these two fossil genera and is the first serpent cranial material recovered from Oligocene strata of North America. It is also the only known specimen in the world in which more than one complete snake is preserved and it suggests that aggregation, a behaviour that is extremely common in living snakes, was already a feature of serpent social behaviour at least 32 million years ago.

rely on 'back-selection' to purge resistance genes from the population after application of the pesticide was discontinued.

If dosage levels, migration, refugia, natural enemies and other factors are to be managed so as to slow down the evolution of pesticide resistance, coordination of efforts over large regions will be required. Some crops lend themselves to this, and some do not. Often the best interests of individuals will differ from those of groups, leading to problems that are social and

political rather than purely biological. Beyond this, even with good biological understanding and coherent planning of group actions, I think it will often be that different sectors — pesticide manufacturers, farmers, and the planners responsible for feeding people — have different aims, stemming from different rates of discounting the future and the absence of a truly common coinage. Population biology can make these tensions clearer, but it cannot resolve them. □

1. Crow, J.F. & Kimura, M. *An Introduction to the Theory of Population Genetics* (Harper and Row, New York, 1970).
2. Mayr, E. *Animal Species and Evolution* (Harvard University Press, 1963).
3. Endler, J.A. *Geographic Variation, Speciation and Clines* (Princeton University Press, 1977).
4. Comins, H.N. *J. theor. Biol.* **64**, 177 (1977).
5. Comins, H.N. *J. theor. Biol.* **77**, 171 (1979).
6. Georgiadi, G.P. & Taylor, C.E. *J. econ. Ent.* **70**, 319 (1977).
7. Taylor, C.E. & Georgiadi, G.P. *Environ. Ent.* **11**, 746 (1982).
8. Hassell, M.P., Lawton, J.H. & May, R.M. *J. anim. Ecol.* **45**, 471 (1976).
9. Stubbs, M. *J. anim. Ecol.* **46**, 677 (1977).
10. Bellows, T.S. *J. anim. Ecol.* **50**, 139 (1981).
11. Tabashnik, B.E. & Croft, B.A. *Environ. Ent.* **11**, 1137 (1982).

12. Morse, J.G. & Croft, B.A. *Entomophaga* **26**, 191 (1981).
13. Croft, B.A. & Brown, W.A. *A. Rev. Ent.* **20**, 285 (1975).
14. Mullin, C.A. *et al. Science* **217**, 1270 (1982).
15. Rock, G.C. *J. econ. Ent.* **72**, 293 (1979).
16. Roush, R.T. & Plapp, F.W. *J. Econ. Ent.* **75**, 304 (1982).
17. Plapp, F.W. & Vinson, S.B. *Environ. Ent.* **6**, 381 (1977).
18. Roush, R.T. & Hoy, M.A. *J. econ. Ent.* **74**, 142 (1981).
19. Hooper, G.M.S. & Brown, A.W.A. *J. econ. Ent.* **58**, 824 (1965).
20. Wood, R.J. *Parasitology* **82**, 69 (1981).

Robert M. May is Class of 1877 Professor of Zoology, Princeton University, New Jersey 08544, USA.

Growth factors

Neuropeptides as mitogens

from Michael R. Hanley

THE current enthusiasm for the study of proliferative growth factors is mostly based on the conviction that they are mitogens, endogenous regulators of cell division. A favourite model for mitogenic signalling is experimentally induced cell division, where both quiescent and normally dividing cell populations are stimulated to divide rapidly and replace damaged tissue. The stimuli for tissue repair are unknown but most speculation has focused on tissue-derived or blood-born mitogens such as platelet-derived growth factor. On page 61 of this issue, however, Nilsson *et al.* report that two related neuropeptides contained in sensory neurones, substance P and substance K, can act as mitogens for connective tissue¹. This discovery is particularly significant because both peptides may also participate in other aspects of the reaction to injury: sensory signalling and neurogenic inflammation². The discovery has important implications for other peptides found in peripheral nerves and also suggests that neural factors have a previously unsuspected role in the coordination of cell proliferation.

The authors use the standard strategy for identifying potential mitogens; the target cells are isolated in primary culture and the onset of cell division monitored autoradiographically by measuring the incorporation of ³H-thymidine into nuclei after stimulation with the potential mitogen. This permits quantification of responses in a morphologically defined cell population and provides a selective measure of a rapid initial event in mitosis, entry into DNA synthesis. In two connective tissue populations, skin fibroblasts and vascular smooth muscle cells, Nilsson *et al.* show a potent, dose-dependent stimulation of DNA synthesis in the presence of the tachykinin neuropeptides substance P or substance K, but not to another neuropeptide, bombesin.

The actions of substance P and substance K seem to be mediated by an authentic tachykinin receptor because their effects are reduced by a specific tachykinin antagonist. Recently, the existence of multiple tachykinin receptors, which may respond to different endogenous peptides, has been suggested³. Substance K has a greater potency than substance P, suggesting that a specific subtype of tachykinin receptor³ is involved in the mitogenic responses, which poses the interesting possibility that the mitogenic effects of tachykinins may be associated uniquely with this class of receptors, and that other cell populations bearing these receptors may also have proliferative responses to tachykinins.

The work of Nilsson *et al.*¹ is the latest contribution to the accumulating evidence that neuropeptides can act as mitogens. For

some time, Rozengurt and his colleagues have explored interactions between mitogenic signals in a cell-line model of quiescent cells, serum-starved 3T3 cells. In this system, they have correlated mitogenic activity with the peptides vasopressin⁴ and bombesin⁵. Unlike substance P and substance K, however, these peptides have a complicated, possibly synergistic, interaction with other mitogens in 3T3 cells. This difference reflects fundamental differences in the responses of a growth-arrested permanent cell line and a normally quiescent primary culture cell. Nevertheless, the 3T3 model provides a simple prediction that vasopressin and bombesin may be mitogenic for normal cells. Vasopressin has been reported to stimulate DNA synthesis in chondrocytes⁶ and bone marrow cells⁷; bombesin, or its mammalian relative gastrin-releasing peptide, can be a growth factor for bronchial epithelial cells⁸. As with the tachykinins, the possibility that these peptides may be endogenous mitogens gains added interest because related peptides occur in mammalian peripheral nerves^{9,10}, providing ideal local sources.

The tachykinin-stimulated DNA synthesis is neither additive nor synergistic with that elicited by platelet-derived growth factor. In other systems, this observation has been taken as evidence that a common effector mechanism is involved, but it is pertinent to consider whether this explanation is plausible. Elsewhere, the inositol phospholipid pathway, recently delineated in some detail, has been proposed as a common route in the actions of all mitogens¹¹. Significantly, the tachykinins, vasopressin and bombesin all stimulate the inositol phospholipid pathway and mobilize cellular calcium, possibly as a direct result of producing the second messenger inositol 1,4,5-trisphosphate, which releases calcium from a non-mitochondrial intracellular pool¹¹. Consequently, any receptor acting on the inositol phospholipid pathway may ultimately stimulate cell division. Whether this is a useful generalization for peptides awaits a detailed evaluation of the inositol phospholipid-regulating peptides for their mitogenic activity on target cells. If, as already seems likely, only a subset of such peptide receptors has mitogenic effects, it will be a major challenge to identify any unique aspects of the biochemistry of lipids and second messengers that may correlate with the stimulation of DNA synthesis.

The next logical step will be to evaluate the mitogenic potential of the tachykinins *in vivo*. There are already some preliminary pieces of evidence that strongly support the idea that sensory neurones containing these peptides participate in both normal and stimulated cell growth. A single neonatal

dose of capsaicin, the pungent agent of capsaicum peppers, permanently destroys a population of sensory neurones that are involved with pain perception and nerve-mediated inflammation¹². Although this population is chemically heterogeneous, it includes the tachykinin-containing neurones, so neonatal capsaicin treatment provides a method of deleting the cellular influence of tachykinin neurones. Intriguingly, lesions in skin¹² and corneal epithelium¹³ frequently result from this treatment. The mechanism of the lesion formation has not been examined in detail but morphological studies suggest that in the cornea the proliferating basal layer of epithelial cells and the normal geometry of cell renewal are severely impaired¹². Clearly, a specific role for the tachykinins cannot be inferred, but this evidence suggests that a factor from sensory neurones is essential for normal cell turnover in the cornea and, possibly, in the skin.

If corneal damage were to result from the loss of a mitogenic peptide input, could there be a corresponding disorder arising from the local overproduction of a mitogenic peptide? Although skin keratinocytes have not been shown to be targets for tachykinins, there is a large density of tachykinin nerves, presumably of sensory origin, in the basal layers of the epidermis¹⁴, which contains actively mitotic cells that divide to replace the outer layers of skin. Local factors are presumed to regulate keratinocyte proliferation, but little is known about such factors. Psoriasis is a skin disorder in which normal control is lost and the cells proliferate at a greatly increased rate. An explanation for the condition may be that the sensory tachykinins exert a tonic control over normal basal layer cell division, and that, in psoriasis, there is an excessive mitogenic stimulation of keratinocytes. This hypothesis could be conveniently evaluated by the topical use of tachykinin antagonists.



100 Years Ago

REPORTS from Japan state that grave fears were entertained of an outbreak of the long quiescent volcano Fujiyama, and that officials had been sent to investigate the matter. The people living in the neighbourhood believed an eruption to be imminent, because, while the snow on the mountain had begun to melt two months before the usual time, all the wells at the fort became dry, and difficulty was experienced in procuring water. The phenomenon is considered the more remarkable from the fact that the winter has been unusually cold, and that the surface of the snow remains hard, the part nearest the ground being the first to give way. Intelligence has also been received from Java of the eruption of the Semiroo mountain, the most active Javanese volcano. From *Nature* 31 610, 30 April 1885.

Strictly, the conclusions of Nilsson *et al.* apply only to peptides found in sensory neurones. However, the results with bombesin and vasopressin^{4,5}, which have not so far been found in sensory neurones, strongly suggest that other parts of the peripheral nervous system may also exert proliferative influences on normal, stimulated or developing cell populations. One intriguing consequence of these investigations may be to reveal for the first time a defined cellular origin for an identified substance with mitogenic action. □

Michael R. Hanley is in the Department of Biochemistry, Imperial College of Science and Technology, London SW7 2AZ, UK.

Protein structure

Hands on the calcium switch

from C. Schutt

THERE is no such thing as a boring protein; newly solved protein structures always contain surprises. Yet the calmodulin structure published on page 37 of this issue¹, and the closely similar troponin-C molecule^{2,3}, turn out to be even more interesting than usual. The simple 'dumbbell'-shaped molecules, with their easily grasped polypeptide fold, invite close scrutiny of the fundamental mechanism of molecular switching.

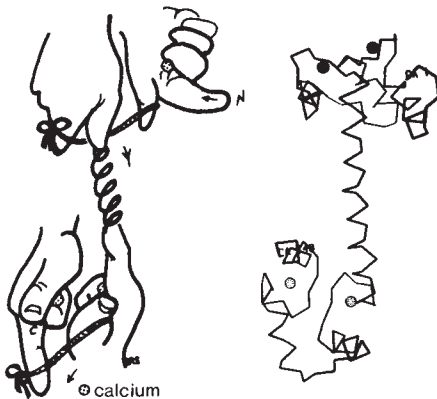
These proteins sense intracellular calcium signals and trigger, through changes in conformation, appropriate responses in associated proteins. Calmodulin is a ubiquitous, highly conserved, eukaryotic protein implicated in many coordinated, calcium-dependent cellular responses⁴, whereas troponin-C is part of a protein complex bound to actin filaments in vertebrate muscle that 'switches on' contraction when calcium is released from the sarcoplasmic reticulum⁵.

The three-dimensional organization of calmodulin is easy to visualize: knobs at both ends of a long exposed helix, with each knob having paired, back-to-back, calcium-binding subdomains (EF-hands) of the type first described for parvalbumin⁶. An EF-hand is a structural motif (see figure) with a 12-residue calcium-binding loop, pictured as the middle finger of the right hand, flanked by two nearly perpendicular helices, represented by thumb and forefinger. In the crystal structure, the concluding helix of the second calcium-binding subdomain runs without interruption into the starting helix of the third binding subdomain, forming the long central helix. It is the splendid isolation of this helix that is unexpected — the tertiary fold of the knobs had been predicted.

In a pioneering use of sequence homology to predict protein structure, troponin-C was determined to have four calcium-binding subdomains⁷⁻⁹. Using the criterion of compactness and structural

1. Nilsson, J., Von Euler, A.M. & Dalsgaard, C.J. *Nature* **315**, 61 (1985).
2. Jessel, T.M. & Womack, M.D. *Trends Neurosci.* **8**, 43 (1985).
3. Lee, C.M., Iversen, L.L., Hanley, M.R. & Sandberg, B.E.B. *Naunyn-Schmiedeberg's Arch. Pharmac.* **318**, 281 (1982).
4. Rozengurt, E., Legg, A. & Pettican, P. *Proc. natn. Acad. Sci. U.S.A.* **76**, 1284 (1979).
5. Rozengurt, E. & Sinnett-Smith, J. *Proc. natn. Acad. Sci. U.S.A.* **80**, 2936 (1983).
6. Miller, R.P., Husain, F., Svenssen, M. & Lohin, S. *Endocrinology* **100**, 1365 (1977).
7. Hunt, N.H., Perris, A.D. & Sandford, P.A. *J. Endocr.* **72**, 5 (1977).
8. Willey, J.C., Lechner, J.F. & Harris, C.C. *Expl Cell Res.* **153**, 245 (1984).
9. Hanley, M.R. *et al.* *Nature* **309**, 258 (1984).
10. Moghizadeh, F. *et al.* *Neuroscience* **10**, 553 (1983).
11. Berridge, M.J. *Biochem. J.* **220**, 345 (1984).
12. Buck, S.H. & Burks, T.F. *Trends Pharmac. Sci.* **4**, 84 (1983).
13. Fujita *et al.* *Expl Eye Res.* **38**, 165 (1984).
14. Weihe, E., Hartschuh, W. & Reinecke, E. in *Substance P*. Dublin 1983 (eds Skrabanek, P. & Powell, D.) 143 (Boole Press, Dublin).

homology with the parvalbumin molecule, Kretsinger and Barry produced a pleasing model of troponin-C with four EF-hands symmetrically packed around a hydrophobic core¹⁰. This model has recently been challenged by Tsalkova and Privalov¹¹, who argue from thermal denaturation results that the protein melts as if comprising three independently folded 'cooperative units'. They conclude that the middle part of the sequence, now



known to form the central helix, somehow prevents the other two units from interacting — a remarkable prediction in view of the three-dimensional structures just published. It is also known that limited proteolysis can split these molecules into two intact but non-associating calcium-binding domains¹². Would this information have helped to get the prediction right? Perhaps such zero-dimensional experiments will become a more important component in future attempts to predict protein structure from sequence homology.

How might the binding of calcium to such a structure act as a conformational switch? In the case of troponin-C, two of the calcium-binding loops (in the amino-terminal lobe) lack calcium in the crystal and, significantly, have altered inter-helix angles from the standard EF-hand, the

thumb and forefinger being nearly parallel. These differences are large enough to be recognized by other proteins, especially if there is a relative rotation of the two lobes about the axis of the central helix. A more radical (but not unreasonable) suggestion by Herzberg and James² is that a glycine residue in the middle of the long central helix serves as a pivot about which the two halves of the dumbbell can freely rotate, perhaps even coming into contact. Such a dramatic change could certainly serve as a switch, possibly initiated by ligand-induced changes in the calcium-binding subdomain at the amino-terminus of the helix.

The question, then, is whether the secret of this molecular switch lies with the instability of the long bridging helix? Proteolytic cleavage studies show that this region does have an altered conformation in the absence of calcium. Furthermore, thermal denaturation studies¹¹ identify the helix as the least stable part of the structure. Perhaps an exposed helix — a rare feature in protein structures — tottering on the edge of conformational stability has exactly the required sensitivity to respond to energies produced by calcium binding. One is reminded of glucagon, a small protein of borderline stability in solution, which only reveals its helical conformation in the crystalline state¹³. Such a delicate balance between order and disorder could be the basis of switching in all these molecules.

It is tempting to carry these speculations too far. Calmodulin and troponin-C are only functional while interacting with other proteins and the central helix could well be buried in these situations, or more subtle aspects of protein-protein recognition might predominate. Nevertheless, particularly by the use of genetically engineered mutants, these proteins may yield deep insights into the relationship between the signalling and stability properties of proteins. Finally, we might ask why such a simple structure was not anticipated, when most of its subdomains had been correctly predicted? My guess is that someone, somewhere, did get it right — but who would want to go down as the first to predict a dumbbell when, if the prediction was wrong, their name might so easily become attached to it? □

1. Babu, Y.S. *et al.* *Nature* **315**, 37 (1985).
2. Herzberg, O. & James, M.N.G. *Nature* **313**, 653 (1985).
3. Sundaralingam, M. *et al.* *Science* **227**, 945 (1985).
4. Cheung, W.Y. *Science* **207**, 19 (1980).
5. Weber, A. & Murray, J.M. *Physiol. Rev.* **53**, 612 (1973).
6. Kretsinger, R.H. & Nockolds, C.E. *J. biol. Chem.* **248**, 313 (1973).
7. Kretsinger, R.H. *Nature new Biol.* **240**, 85 (1972).
8. Collins, J.H., Potter, J.H., Horn, M.J., Wiltshire, G. & Jackman N. *FEBS Lett.* **36**, 268 (1972).
9. Tufty, R.M. & Kretsinger, R.H. *Science* **187**, 167 (1975).
10. Kretsinger, R.H. & Barry, C.D. *Biochim. biophys. Acta* **405**, 40 (1975).
11. Tsalkova, T.N. & Privalov, P.L. *J. molec. Biol.* **181**, 533 (1985).
12. Newton, D.L., Oldewurtel, M.D., Krinks, M.H., Shiloach, J. & Klee, C.B. *J. biol. Chem.* **259**, 4419 (1984).
13. Panjipan, B. & Gratzner, W.B. *Eur. J. Biochem.* **54**, 547 (1974).

C. Schutt is at the Medical Research Council Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK.

Environmental science

Acidification and decline of Central European forests

from Nico van Breemen

WHAT is probably the most complete set of data available on the effects of atmospheric sulphur and nitrogen on the weathering of soil and rock, and on the composition of drainage water in forested and agricultural areas, is reported on page 31 of this issue by Thomas Paces¹. Of particular interest is that one of the areas (all of which are in Czechoslovakia) is characterized by extremely high sulphur inputs and a complete dieback of spruce forest, a situation that may be typical of several industrial regions in eastern Europe. Paces concludes that the forest in the industrial region is strongly acidified. But is acidification responsible for the dieback?

Dry deposition of sulphur dioxide is responsible for 75 per cent of the acidification in the industrial region whereas wet deposition (acid rain in the strict sense) contributes only 15 per cent to the total acid load of the system studied. In relatively remote forests, the acid load is only about a third of that in the industrial region, and the contributions of wet and dry deposition to the total acid load are about equal. Soil acidification is considerable in such areas but the drainage water is almost neutral. Agricultural activities (fertilizer application and harvesting) contribute considerably to the acid load of soil and rock, resulting in increased mineral weathering.

Changes in the chemical composition of water of the Elbe river over the past 80 years indicate that these conclusions, although based on data drawn from four small representative watersheds, are valid for large areas. All the four have similar bedrocks, soils and climates but differ in their exposure to atmospheric pollution and land use. Paces combines detailed hydrological measurements and estimates of chemical input and output (including records of fertilization and harvesting) over five to seven hydrological years with a recently developed steady-state model of weathering, to quantify acid production and consumption by known geochemical and biological processes. Inputs of sulphur and nitrogen by dry deposition were estimated from measured concentrations and published values of deposition velocities.

Although some authors still contest that atmospheric deposition of anthropogenic sulphur and nitrogen seriously contribute to acidification of soils^{2,3}, Paces's data are in agreement with several independent lines of evidence for increased environmental acidification as a result of atmospheric deposition. The first direct evidence for the strong soil acidification in Central Europe of the past decades came from pH measurements of old and new soil samples by

Butzke⁴. At a workshop at Uppsala last September, C.O. Tamm from the University of Uppsala reported a small drop in the pH of forest soils over the past 50 years in northern Sweden but a much larger acidification in the southern part of the country. Tamm attributes the small pH decrease in the north mainly to biological processes and the large decrease in the south to atmospheric inputs. Other input-output budgets⁵, geochemical studies of soil weathering profiles⁶ and simulation models of soil acidification⁷ also lead to the same inescapable conclusion: the rate of acidification of forest soils in a large part of Europe (and north-eastern America) is greatly accelerated by atmospheric pollution. Although this acidification has markedly depressed the pH of sensitive forest soils in Central Europe, forest soils in the United States and Canada seem still to be able to buffer the pH in a range that can be considered normal.

Although there is general agreement that the acidification of lakes and streams is detrimental to fish and other biota⁸, the relation between soil acidification and the widespread decline in forests of Central Europe is still controversial; the link between the acidification of soil and death of spruce forests indicated by Paces is not proven. Usually, direct effects of atmospheric gases, including photo-oxidants, may be more important than soil acidification in damaging trees. Nevertheless, the removal of base cations and release of potentially toxic aluminum by acidification must be detrimental to trees and other forest plants.

In the long run, soil acidification may be a bigger problem than the direct effects of gases on vegetation: although a reduction in pollutant emission would have an almost instantaneous effect on air quality, recovery of acidified soils is presumably very slow and is only possible if sufficient quantities of weatherable silicate minerals are still present.

In that respect, the acidification of agricultural soils (also demonstrated by Paces) is a minor problem. The acidifying effects of harvesting and fertilizer application are normally counteracted by liming. Additional effects of atmospheric acids can easily be undone in the same way.

Liming, however, probably offers no practical solution for forests. Apart from the fact that acid deposition rates are higher in forests than on land⁹ and the problem of how lime could be applied in forest areas, forest liming could result in undesirable side effects from increased decomposition of organic matter, nutrient imbalances, increased susceptibility to pests and diseases and changes in the species composition of the ground vegetation. Both the direct effects of atmospheric pollution and the long-term problems associated with acidification of the environment can only be solved by a reduction in the emissions of gaseous sulphur, nitrogen and hydrocarbons. □

1. Paces, T. *Nature* **315**, 31 (1985).2. Krug, E.C. *Nature* **313**, 73 (1985).3. Tabatabai, M.A. *CRC Crit. Rev. Envir. Control* **15**, 65 (1985).4. Butzke, H. *Forst u. Holzwirt* **36**, 542 (1981).5. van Breemen, N., Driscoll, C.T. & Mulder, J. *Nature* **307**, 599 (1984).6. Mazzarino, M.J., Heinrichs, H. & Fölster, H. in *Effects of Accumulation of Air Pollutants in Forest Ecosystems* (eds Ulrich, B. & Pankrath, J.) 113 (Reidel, Dordrecht, 1983).7. Kauppi, P., Kämari, J., Posch, M., Kauppi, L. & Matzner, E. *Europ. Ecol. Modelling* (in the press).8. Muniz, I.P. *Phil. Trans. R. Soc. B* **305**, 517 (1984).9. Ulrich, B. & Pankrath, J. (eds) *Effects of Accumulation of Air Pollutants in Forest Ecosystems* (Reidel, Dordrecht, 1983).

Nico van Breemen is in the Department of Soil Science and Geology, Agricultural University, 6700 AA Wageningen, The Netherlands.

Primate cognition

Can apes learn to count?

from Brendan McGonigle

DANZIGER once described mathematics as the language of size, natural language as the language of sort. Are these languages confined to the human brain? Although research (and controversy) continues on whether apes share important features of our natural language system, attention is now turning to the question of whether they might also possess the rudiments of a number system. Thus, on page 57 of this issue, Tetsuro Matsuzawa reports on the 'use of numbers' by a chimpanzee¹.

Matsuzawa claims that the ape successfully identified six arabic numbers by selecting from a keyboard the numeral that was appropriate to the number of objects displayed. In addition, it provided

the appropriate colour name and category label for some 300 sample types, stringing these into, for example, '3 red pens'.

The study is perhaps unique in requiring a level of description of an array of objects in terms of both numerosity and the type of object displayed. These findings are surely provocative, the fairground appeal of an ape that selects numerals felicitously apart. Do they, however, imply that the chimpanzee is endowed with the rudiments of a system of calculation which finds a refined and rarified expression in the axiomatic systems devised by man? Only a convergent programme of comparative developmental research into the natural or intuitive calculus of animals and humans

will answer this question. Meanwhile, it is instructive to evaluate these and related findings in the light of current advances in child development, which already implicate some basic psychological mechanisms underlying mathematical understanding, particularly the conception of number.

Subitizing — the simultaneous apprehension of up to four elements through direct pattern recognition — is one rudimentary counting mechanism that has been identified in humans and is, perhaps, implied in the chimpanzee by Matsuzawa's study. The process is revealed by the measurement of the time it takes subjects to enumerate small numbers of items, such as dots, on a visual display. Increasing the number of items up to four has relatively little effect, adding around 200 milliseconds per extra dot with child subjects; above that number, however, each additional dot increases the enumeration time by over 1000 milliseconds. This 'break' in the range of times is consistent with the notion that subitizing, or simultaneous number perception, applies for up to four items, beyond which counting, a more costly procedure, begins.

It is impossible to evaluate the role of subitizing in Matsuzawa's study as identification measures are not reported. It is, however, interesting that the chimpanzee made by far the greatest number of mistakes (over 63 per cent) in the discrimination between numbers five and six, just outside the subitizing range for humans. It may be that the span of apprehension for apes is different from that of man. This needs to be determined empirically, and the research could also clarify the contentious relationship between subitizing and counting in the genesis of the number system².

Subitizing apart, developmentalists are agreed that even 'true' counting begins with small numbers. Many six-year-olds cannot count quantities above ten. Initially, there is an apparent need to use '1' as an anchor; asked to begin counting from a higher number, performance falters. And forward counting begins before counting backwards, which indicates that there is a preliminary temporal or serial-order basis for the number system. (It is interesting to note that temporal succession was suggested by both the philosopher Kant and the mathematician Hamilton as an intuitive basis for number.) The subsequent development of successor links in the counting chain seems to depend largely on a mental representation with (spatial) analogue properties³. This is inferred from the inverse relationship between the time taken to identify the greater of two numbers and their ordinal separation (for example, it takes longer to decide that $5 > 4$ than $7 > 4$) and suggests that there is a 'mental line' or scale on which numbers are mapped.

Such phenomena are seen to reflect important stages in the development of informal arithmetic skills, so their investi-

gation in animals would help inform us whether a rudimentary calculus is shared with other species. Already, we have reported evidence for a representational device analogous to the 'mental line' (also based on decision-time data) for squirrel monkeys tested on a related symbolic (logic) task⁴. The expectation is therefore provoked that comparative research into number representation could reveal the communality of at least some counting mechanisms in humans and other primates.

The use of number, however, implies more than mere counting skill *per se*. Children count before applying their knowledge in situations where it would be desirable to do so. Asked to determine which of two rows of objects has 'more', four-year-old children tend to base their judgments on the shape of the set, selecting the more extended row and thus appearing

to treat discontinuous quantities as continuous ones⁵. Yet children who fail in these tests can count under other circumstances. Should true counting be found in non-human primates, therefore, it would be necessary to evaluate their spontaneous use of counting as a measuring device, especially under conditions which afford other (if less reliable) solutions. □

1. Matsuzawa, T. *Nature* 315, 57 (1985).
2. Klahr, D. & Wallace, J.G. *Cognitive Development: An Information Processing View* (Erlbaum, New Jersey, 1976).
3. Resnick, L. in *The Development of Mathematical Thinking* (ed. Ginsburg, H.P.) Ch. 3 (Academic, New York, 1983).
4. McGonigle, B.O. & Chalmers, M. in *Reasoning and Discourse* (eds. Myers, T.E., Brown, E.K. & McGonigle, B.O.) Ch. 6 (Academic, London, 1985).
5. Piaget, J. *The Child's Conception of Number* (Routledge and Kegan Paul, London, 1969).

Brendan McGonigle is in the Department of Psychology, University of Edinburgh, 7 George Square, Edinburgh EH8 9JZ, UK.

Biopolymers

The spinning of silk

from Paul Calvert

SILK is stored as a concentrated aqueous solution in glands of the larval silkworm until it is spun as a dry fibre by a gentle waving of the head of the larva. The equivalent process for spinning a nylon fibre — the synthetic fibre that is closest to silk — consists in forcing the molten polymer out through a small jet, a spinneret, and cooling it. The resulting fibre is soft and not very strong, as it is not possible to run the polymer out through the small jet fast enough to align the molecules. Alignment is only achieved in a second process where the fibre is cold-drawn to about six times its original length, whereupon the molecules become oriented parallel to the fibre axis and the fibre becomes strong and stiff. So how does the silkworm manage to produce the same effect in a single step at room temperature? As silk fibres are insoluble in water, the solution stored in the silk gland must be metastable yet remain liquid until the mechanical process of spinning induces precipitation; no chemical change or drying seems to occur during spinning. But how does the apparently slow process of spinning a mobile liquid nonetheless result in a highly ordered fibre? And why does the production not become blocked up with the precipitating solid?

Most of the recent published work on the physiology of silk production has been carried out in Japan and much of it is summarized in a group of reviews in *Experientia* 39, 441; 1983. Silkworm silk is secreted as a 15 per cent aqueous solution of fibroin in the narrow, convoluted posterior gland. From there it moves to the middle storage gland where the fibroin is enveloped in a solution of sericin that will ultimately become the gummy coating holding the fibres together in the cocoon. It is in the middle gland, where the concen-

tration of silk rises to 30 per cent, that one might expect the polymer to be precipitated out. Akai (*Experientia* 39, 443; 1983) describes the fibroin forming fibrous spherical masses of about 1 μ m in diameter, but an electron microscope study of such a system is prone to artefacts and the conformation of the fibroin has been much debated.

In its most stable fibre form, silk has a β -sheet structure with chains extended and packed into layers. If the solution in the silk gland is removed and allowed to dry undisturbed, it forms silk I. Because this can only be obtained as crystal aggregates and not as oriented fibres (Lotz, B. & Colonna-Cesari, F. *Biochimie* 61, 205; 1979), it does not lend itself to a complete X-ray structure determination. Nevertheless, a structure has been suggested in which the conformation of alanine and glycine residues are near the β -sheet and α -helix, respectively, and which is much more contracted than β -silk. But new data from magic-angle ¹³C nuclear magnetic resonance instead suggest a loose 4-fold helix (Saito, H. *et al. Macromolecules* 17, 1405; 1984).

In the liquid itself, the silk has been believed to be predominantly in a randomly coiled state with some α -helix, but another recent ¹³C nuclear magnetic resonance study (Asakura T. *et al. Macromolecules* 17, 1075; 1984) shows that it is randomly coiled in dilute aqueous solution and probably changes over to the loosely helical conformation of silk I as the concentration rises above five per cent, when the chains start to aggregate. This loose helix probably provides a conformational energy minimum that is readily reached and so prevents the chains from going over to the more stable β -sheet form and solidifying. The block copolymer structure of fibroin may help to stabilize the loose helix.

The concentrated liquid is certainly quite fluid: liquid readily flows out of the cut ends of glands of spinning silkworms (Magoshi, J., Magoshi, Y. & Nakamura, S. *Rep. Progr. Polymer Phys., Japan* **23**, 747; 1980). The precipitation step that occurs during spinning seems to depend on a critical shear rate being reached. *In vitro* measurements on solutions of less than five per cent show that precipitation occurs if liquid silk is sheared at rates greater than 50 s^{-1} . The silkworm spins the fibre at about 1 cm s^{-1} from the anterior gland, which narrows to $50\text{ }\mu\text{m}$ at the entrance. This corresponds to a shear rate of about 1 s^{-1} and causes precipitation of silk at the much higher concentrations found *in vivo* (Kataoka, K., & Uematsu, I. *Kobunshi Ronbunshu Engl. Ed.* **6**, 621; 1977). Magoshi *et al.* also report that the change to the β -sheet structures occurs at draw ratios of about four and that the silk can be induced to flow out of cut glands in a liquid crystal form.

From recent efforts to prepare high-strength synthetic fibres, it has become clear that to produce high levels of orientation the inter-chain forces must be

controlled so as to permit slippage of the chains past one another while preventing rapid relaxation back to a randomly coiled state. Strong aromatic polyamide fibres such as Kelvar, are produced by spinning from liquid crystalline solutions in which the chains are locally aligned even before starting, thereby greatly increasing the orientation. Apparently silk production is similar, in that the aggregated state promotes local alignment of the chains while still keeping them in solution. They then easily spin to high orientations and to moduli that are comparable with the highest values achievable with nylons and about half those of Kelvar. The change from coiled to rigid conformation that happens with increasing silk concentration apparently corresponds to the 'induced rigidity', believed to occur in stiff polymers in solution (Matheson, R.R. *Mol. Cryst. Liq. Cryst.* **105**, 315; 1984). Again we have failed to recognize a natural process until rediscovering it the hard way. □

Paul Calvert is in the School of Chemistry and Molecular Sciences, University of Sussex, Falmer, Brighton BN1 9QJ, UK.

Icythyology

The biology of coelacanths

from Jared M. Diamond

ON 22 December 1938, a taxi driver on the docks of East London, South Africa, was hailed by a lady with a very oily dead fish that proved to be the coelacanth *Latimeria chalumnae*, sole survivor of a group of fish that had been believed extinct for 70 million years and traditionally considered close to the ancestor of all land vertebrates. Since then, many other specimens have been captured and pickled or frozen but scientists have yet to observe a healthy live specimen. But much has been learned from preserved specimens about ecology, sensory biology, reproductive biology, physiology, and taxonomic affinities of this 'living fossil'¹⁻³.

As for *Latimeria*'s geographical distribution and habitat, all but the first specimen were caught by native fishermen off two of the Comoro islands, between Madagascar and the east coast of Africa. Most specimens have been taken at depths of 100-300 metres as an accidental by-product of night fishing for oilfish, and the apparent restriction of *Latimeria* to these two islands is probably an artefact of how socio-economic and oceanographic conditions constrain the search for oilfish. The South African specimen taken 2,000 miles from the Comoros is the only proof of a wider distribution, but coelacanths may eventually turn up at other seamounts and banks in the western Indian Ocean.

Most likely, coelacanths perch on a reef, lunge for prey and suck it in, rather like a grouper². They feed on fish and cuttlefish, swallowed whole, that live on or

near the bottom of reefs in deep water. *Latimeria*'s most distinctive anatomical feature relevant to feeding is a joint across the skull that permits rotation by up to 15 degrees and thereby widens the gape and increases the strength of the bite⁴. Such intracranial joints are known from some fossil fish but no other living fish. Sensory equipment for prey detection includes large, nearly colour-blind eyes that are adapted to low light levels, densely packed rods but very few cones, and a large cavity in the snout, the rostral organ, which has the structure of an electroreceptor⁵.

One of the most surprising discoveries about *Latimeria* concerns its reproductive biology. Initially, the coelacanth was assumed to lay eggs for external fertilization, like most fish, because males lack a penis, pelvic claspers, modified anal fins or any other obvious copulatory organ. Then, dissection of a female revealed shell-less and very large eggs, 8.5-9 cm in diameter⁶. Finally, Smith *et al.* proved that the eggs are fertilized and hatch internally by finding five miniature coelacanths, 30-33 cm long, in the oviduct of another female⁷. Dates of capture suggest that the gestation period is around 13 months. How the male achieves internal fertilization remains a mystery; two pairs of erectile caruncles flanking the cloaca may play a part^{8,9}.

The other big surprise concerns the osmoregulation of *Latimeria*. Living marine animals osmoregulate in one of three ways: by being iso-osmotic and

similar in ion concentrations to sea water (hagfish and marine invertebrates); by being iso-osmotic, despite having ion concentrations far below those in sea water, through retaining urea and trimethylamine oxide (sharks); or by having ion concentrations far below those in sea water and being hypotonic but drinking sea water and excreting salt to remain in water balance (modern bony fish). *Latimeria* proves to resemble sharks rather than modern bony fish^{10,11}. With 197 mM Na⁺ and 187 mM Cl⁻ in its blood, as in bony fish, and 377 mM urea and 122 mM trimethylamine oxide, both the blood and urine of the coelacanth are iso-osmotic to sea water.

The osmoregulatory resemblance of *Latimeria* to sharks begs the question of its taxonomic affinities. Traditionally, coelacanths have been considered the sister group of the rhipidistians, an extinct group of bony fish that gave rise to amphibians and thence to other land vertebrates. Lagios, however, considers coelacanths the sister group of chondrichthyans (sharks and rays) on the basis of similarities not only in urea retention but also in pituitary structure, pancreatic structure and the presence of a salt-secreting rectal gland^{2,12}. Other authors argue that these similarities do not necessarily ally coelacanths with sharks^{2,10,13}. For example, urea retention arose independently in lungfish, certain frogs and some primitive ray-finned fish and so may also have arisen independently in coelacanths and sharks.

Thus, the dogma on which the current generation of biologists was reared — that coelacanths are close to the ancestors of tetrapods — is under vigorous debate. But the outcome is unlikely to shake their image as a symbol of long preservation without change. Thus, Time Magazine once branded Richard Nixon as a "coelacanth of American anti-communism", and a British MP ridiculed a fellow MP as a 'coelacanth' on the grounds that from his long silence in that august assembly it was a surprise to find him still alive². Whether these calumnies of coelacanths are taxonomically justified is a question that can be resolved only when political conditions allow fresh specimens to be obtained from the Comoro islands or neighbouring waters. □

1. Thomson, K.S. *Nat. Hist.* **82**, No. 2, 58 (1973).
2. McCosker, J.E. & Lagios, M.D. *The Biology and Physiology of the Living Coelacanth* (Occas. Pap. Calif. Acad. Sci. No. 134, 1979).
3. Schopf, T.J.M. *A. Rev. Earth planet. Sci.* **12**, 245 (1984).
4. Thomson, K.S. *Science* **153**, 999 (1966).
5. Bemis, W.E. & Hetherington, T.E. *Copeia* **467** (1982).
6. Anthony, J. & Millot, J. *C.R. Acad. Sci., Paris* **274**, 1925 (1972).
7. Smith, C.L. *et al. Science* **190**, 1105 (1975).
8. Millot, J. & Anthony, J. *Bull. Mus. nat. Hist. Nat. Ser. 2* **32**, 287 (1960).
9. Griffith, R.W. & Thomson, K.S. *Nature* **242**, 617 (1973).
10. Griffiths, R.W. & Pang, P.K.T. *Occas. Pap. Calif. Acad. Sci. No. 134*, 79 (1979).
11. Griffith, R.W. *Proc. R. Soc. B* **280**, 329 (1980).
12. Lagios, M.D. *Copeia* **942** (1982).
13. Lauder, G.V. *Copeia* **942** (1980).

Jared M. Diamond is at the University of California Medical School, Los Angeles, California 90024, USA.

Destruction of fisheries in Africa's lakes

C.D.N. Barel, R. Dorit, P.H. Greenwood, G. Fryer*, N. Hughes, P.B.N. Jackson, H. Kawanabe, R.H. Lowe-McConnell, M. Nagoshi, A.J. Ribbink, E. Trewavas, F. Witte & K. Yamaoka

The introduction of the Nile perch into Lake Victoria, East Africa, has had disastrous consequences. Fisheries have been not merely damaged but destroyed. Lake Malawi has the world's most species-rich fish fauna and the prospect of introduction of alien species could mean disaster.

LAKE Victoria in East Africa, with an area of some 68,635 km², has a naturally rich and diverse fish fauna and formerly supported important fisheries, based largely on herbivores and detritus feeders of the family Cichlidae¹⁻³. Fisheries based on such short food chains are ecologically the most efficient. Although over-fishing in Lake Victoria has for some time been a cause for concern, fisheries technology having advanced more rapidly than the scientific knowledge needed to devise rational means of exploiting this complex multispecific assemblage, a further serious hazard now faces the fisheries. In the 1950s, a suggestion was made that the Nile perch, *Lates niloticus* (Fig. 1), a piscivorous species that grows to a large size, should be introduced into the lake. Although it was pointed out that simple ecological principles show that such an introduction would be detrimental⁴ (yields of predators can never be as great as those of their prey), the idea was pursued by its supporters and in 1960 the Nile perch found its way into the lake, apparently from ponds into which it had already been introduced. Subsequently this species was deliberately introduced into the lake.

Since then the Nile perch has spread rapidly in the northern parts of Lake Victoria, having reached Kenyan waters by the mid-1960s. It is now abundant in many areas and is spreading south. It has invaded the Tanzanian waters of the lake and, in the early eighties, increased rapidly in numbers in the vicinity of Mwanza, where there are important fisheries based on indigenous species⁵.

Invasion

The consequences of this invasion and of the establishment of the Nile perch are serious. While some of them were predicted, not all could have been anticipated. For example, the Nyanza (formerly Kavirondo) Gulf was the site of important fisheries for an endemic tilapia, *Oreochromis esculentus*, and for other fishes including the abundant haplochromine cichlids. Large quantities of the former supplied Nairobi and other centres of population and the latter, easily preserved and transported, were sold over a wide area. Since the arrival of the Nile perch, almost all the indigenous fishes of commercial importance have not merely declined but have virtually disappeared^{6,7}. These include not only the herbivores but

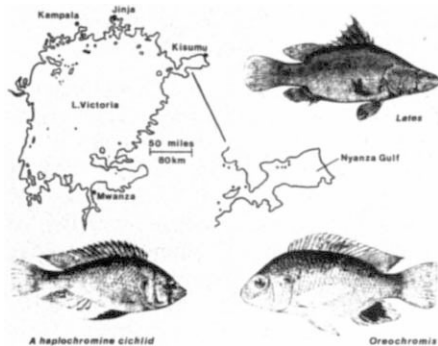


Fig. 1 Lake Victoria and, inset, the Nyanza Gulf, the introduced Nile perch, *Lates niloticus*, and representatives of some of the cichlid fishes, formerly the basis of important fisheries in the Nyanza Gulf and elsewhere, that have recently declined catastrophically. Illustrations of fishes after Boulenger²⁰. *Lates* reach up to 2 m long, *Oreochromis* up to at least 30 cm, and the many species of haplochromine cichlids are usually less than 20 cm in length.

carnivores such as certain catfishes.

Large numbers of *Lates* are being caught, however, reflecting the fact that they now make up well over 80 per cent of the fish biomass in the Nyanza Gulf⁶, an atypical and clearly unsustainable role for a predator at the top of the food chain. When first introduced, *Lates* fed mostly on small haplochromines (Fig. 1), a group estimated formerly to comprise ~80% of the demersal fish stocks of the lake⁸. Having virtually eliminated this food source, small *Lates* now depend on the small prawn, *Caridina*, which is consumed by juveniles in several lakes. Recent observations show that ~80% of the diet of individuals less than 45 cm in length consists of these prawns and even fishes up to 100 cm in length are eating *Caridina*⁶. Large individuals are feeding predominantly on young of their own species.

Such a situation is inherently unstable and the ecology of the Nyanza Gulf has been irreparably upset. Furthermore, *Lates* is not liked by the local people, who constantly complain that they are unable to find their favourite food fishes. Prices are low: at the lakeside, a kilo of *Lates* sells for a shilling, but in areas where it is still

available, a kilo of *Oreochromis* fetches 30 shillings. Moreover, because of its large size and oily nature, *Lates* cannot be sun-dried but has to be smoked, which is leading to the deforestation of a number of islands as trees are felled to produce firewood. The establishment of *Lates* has also had unfortunate sociological consequences. Fishermen who could formerly subsist on small-scale haplochromine and *Oreochromis* fisheries are being driven out of business and production is being concentrated in the hands of wealthier individuals who can afford the larger, stronger and more expensive nets required for *Lates*.

Fishery decline

Against this background, the conclusions of a symposium of the UN Food and Agriculture Organization (FAO)⁹, held in October 1984, make curious reading. The symposium was convened under the auspices of the Committee for Inland Fisheries of Africa as the concern of its subcommittee for the development and management of the fisheries of Lake Victoria. Its purpose was to consider the status of the fish stocks of Lake Victoria, to suggest management measures, and to define priorities for research. Its report confirms that the fisheries, formerly based on a wide range of species or species groups, are now dominated by only four such groups, one of which, that comprising the haplochromine cichlids, "is itself threatened", and that fisheries for most species are in decline. The ill-advised trawl fishery for haplochromines was already in decline before the upsurge of the Nile perch and the report now admits that it is "almost certain" that haplochromine stocks cannot support the twin stresses of heavy predation by *Lates* and a fishery, and thinks it would be "extremely unwise" to invest further in this industrial fishery. In the Nyanza Gulf at least, the haplochromine fishery is already virtually nonexistent. It is admitted that, under what is seen as the best of two scenarios, production from Lake Victoria will settle down at a level "assumed to be some 80% less than the productivity achieved by the

Authors' addresses: Zoologisch Laboratorium der Rijksuniversiteit Leiden, Kaiserstraat 63, 2300 RA Leiden, The Netherlands (C.D.N.B.). Museum of Comparative Zoology, Harvard University, Cambridge, Massachusetts 02138, USA (R.D.). British Museum (Nat. Hist.), Cromwell Road, London SW7 5BD, UK (P.H.G., E.T.). Freshwater Biological Association, The Ferry House, Ambleside, Cumbria LA22 0LP, UK (G.F.). Alaska Cooperative Fisheries Research Unit, University of Alaska, Fairbanks 99701, Alaska, USA (N.H.). J.L.B. Smith Institute of Ichthyology, Post Bag 1015, Grahamstown, S. Africa (P.B.N.J., A.J.R.). Department of Zoology, Faculty of Science, Kyoto University, Kyoto 606, Japan (H.K.). Streatwick, Streat, Hassocks, Sussex (R.H. L.-M.). Faculty of Fisheries, Mie University, 2-80 Edobashi, Tsu, 514 MZ Japan (M.N.). Haplochromis Ecology Survey Team, Zoologisch Laboratorium der Rijksuniversiteit Leiden, Kaiserstraat 63, 2300 RA Leiden, The Netherlands (F.W.). Department of Fisheries, Faculty of Agriculture, Kyoto University, Kyoto 606, Japan (K.Y.).

*To whom correspondence should be addressed.

pre-Nile perch community due to energy losses by the predator". Alternatively, in what is described as a boom or bust situation — another guess — the disrupting influence of *Lates*, which is admitted, is likely to lead to even lower productivity. Despite this it is claimed that "the introduction of Nile perch to utilize the haplochromine cichlids of the lake may be regarded as successful in terms of its original objectives"! With such criteria for success all things are possible. It is also claimed that losses in productivity "may, however, be compensated for by the greater value and quality of the fish caught". Lakeshore prices cast serious doubts on this prediction and no improvement of quality could compensate for the enormous losses in both yields and diversity brought about by the introduction of the Nile perch.

The establishment of *Lates* is not only an economic and ecological tragedy but also an enormous loss to evolutionary biology. Lake Victoria is the site of a remarkable radiation of cichlid fishes which has given rise to several hundred endemic species^{10,11}. As an example of explosive speciation this assemblage is of immense scientific interest¹² and the loss of these fishes, often before their biology has been elucidated, is incalculable.

Irreversibility

We also wish to stress the irreversibility of introductions into large lakes — removal of *Lates* from Lake Victoria is now impossible — and to emphasize that the introduction of this fish is seen as a failure, not merely by a small group of concerned scientists who think they know best and who find the endemic faunas of African lakes of academic interest, but by the people it was meant to benefit. This is no minor incident but a disaster whose magnitude is put into perspective by noting that Lake Victoria is much larger than Switzerland.

Great damage has already been done to Lake Victoria, so it is disturbing to learn that the introduction into Lake Kivu of *Boulengerochromis microlepis*, the largest piscivorous cichlid fish of Lake Tanganyika, is now contemplated. Even more disturbing is the suggestion that the endemic pelagic zooplankton-feeding clupeid fishes of Lake Tanganyika should be introduced into Lake Malawi (= Nyasa). One of these, *Limnothrissa miodon* (Fig.2), has been successfully transferred to, and established in, the man-made Lake Kariba, and into Lake Kivu, a lake of recent origin with an extremely depauperate indigenous fish fauna, and which drains into Lake Tanganyika. These situations are utterly different from those prevailing in Lake Malawi. Malawi is an ancient lake in which evolution has given rise to exceedingly rich assemblages of endemic fishes, whose differentiation has proceeded further than in Lake Victoria, and of which it has more than any other lake in the world. These include a group of zooplankton-feeding

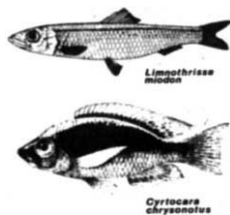


Fig.2 The zooplankton-feeding Tanganyikan clupeid *Limnothrissa miodon* (top, after Poll²¹) and one of the Utaka group of haplochromine cichlids (bottom, after Boulenger²⁰) that occupy a similar niche in Lake Malawi.

haplochromines, the Utaka, that are caught for human consumption. Other zooplankton-feeders are also present, so the niche that would-be introducers wish to exploit is already filled. The great age of Lake Malawi has enabled its fishes to evolve into complex communities whose members interact in many subtle ways, and the effect of the introduction of alien species is impossible to predict. It is unlikely that a situation which has evolved over perhaps one to two million years can be improved by the introduction of an alien species whose impact can only be guessed at. The consequences could be disastrous and irreversible, as they have been in Lake Victoria, in the Laurentian Great Lakes of North America¹³, Lake Titicaca (S. America)¹⁴, Lake Lanao (Philippines)¹⁵ and elsewhere. Similar concern has recently been expressed about introductions into Lake Kinneret, Israel¹⁶. The introduction of seemingly innocuous organisms, even those whose ecology has been much studied, can have unforeseen (and unforeseeable) consequences. Such have followed the introduction of the mysid crustacean *Mysis relicta* into certain lakes where it was hoped it would increase fish yields. For example, in one well-documented case it virtually eliminated planktonic cladocerans, an important food for fishes¹⁷. As a result a group of experts has unanimously approved a resolution that a complete moratorium be imposed on future introductions¹⁸.

With one disaster accomplished in Africa, we urge those interested in improving the fisheries of Lake Malawi, whether in FAO or other bodies offering technical and scientific assistance, to concentrate on rational exploitation of the indigenous fish stocks, efficient distribution of catches and the reduction of unnecessary losses during preservation and storage. It is in addressing these questions, rather than in indulging in potentially dangerous introductions, that true progress can be made.

Lake Malawi is not only an important source of protein but is a site of international significance as a laboratory of evolution. Its future to the welfare of those who crop it, and to the international scientific community, is too important to be jeopardized by those who, with no means of predicting the consequences, would introduce alien fishes. We, of several nationalities, with experience of African

lakes, therefore urge the governments of the states bordering Lake Malawi (one of which is Tanzania whose Lake Victoria fisheries are now threatened by the Nile perch) never to allow such introductions. The introduction of alien species could seriously jeopardize the food supplies and livelihoods of future generations. Such an action, the outcome of which is unpredictable, and scientifically indefensible, could prove disastrous to the continued existence of the viable Malawian fishery.

Conclusions

Fishes have often been moved from lake to lake, sometimes with the laudable goal of increasing the yield of human food. Except in fishless lakes and man-made reservoirs, such efforts have failed to achieve the desired objective, and sometimes the results have been disastrous. When large lakes whose fisheries are of immense importance as sources of human food are involved, especially in areas where other animal protein is scarce, much is at stake. A major scientific objection to such introductions is that the outcome is often unpredictable. Detrimental consequences can sometimes be predicted, as in the case reported here, but as controlled experiments cannot be carried out, and the reactions of the organisms involved cannot be determined in advance, a large element of guesswork is inevitable. The outcome is also often irreversible. An additional danger, not considered here, is that introductions may lead to the introgression of new genes. This can hamper taxonomic and evolutionary studies and hinder progress in aquaculture, as is happening among the economically important tilapia cichlid fishes of Africa¹⁹. □

1. Beverton, R.J.H. *Report on the State of the Lake Victoria Fisheries* (Fisheries Laboratories, Lowestoft, 1959).
2. Garrod, D.J. E. *Afr. agric. For. J.* **27**, 69-76 (1961).
3. Fryer, G. & Iles, T.D. *The Cichlid Fishes of the Great Lakes of Africa: Their Biology and Evolution* (Oliver and Boyd, Edinburgh, 1972).
4. Fryer, G. E. *Afr. agric. J.* **25**, 267-270 (1960).
5. Goudswaard, P.C. & Witte, F. *Rep. Haplochromis Ecol. Survey Team No.32* (1984).
6. Hughes, N.F. *Report of the Oxford University Nile Perch Project 1983* (Oxford University, 1983).
7. Okenwa, E.N. *Hydrobiologia* **108**, 121-126 (1984).
8. Kudhongania, A.W. & Cordone, A.J. *Afr. J. trop. Hydrobiol. Fish.* **3**, 15-31 (1974).
9. *Report of the Symposium on the Main Management and Development Issues Facing the Riparian States of Lake Victoria* FAO CIFA DM/LV/84/7 (1984).
10. Greenwood, P.H. *The Haplochromine Fishes of East African Lakes* (Kraus, Munich, 1981).
11. Oijen, M.J.P., van Witte, F. & Witte-Mass, E.L.M. *Neth. J. Zool.* **31**, 149-174 (1981).
12. Greenwood, P.H. *Bull. Br. Mus. (nat. Hist.) Zool. Suppl.* **6**, 1-13 (1974).
13. Baldwin, N.S. *Can. Audubon Mag.* Nov.-Dec. 2-7 (1964).
14. Villwock, W. *Verh. int. Verein. theor. angew. Limnol.* **18**, 1227-1234 (1972).
15. Frey, D.G. *Verh. int. Verein. theor. angew. Limnol.* **17**, 1090-1102 (1969).
16. Gophen, M., Drenner, R.W. & Vinyard, G.G. *Fish. Mgmt* **14**, 43-45 (1983).
17. Threlkeld, S.T., Rybock, J.T., Morgan, M.D., Folt, C.L. & Goldman, C.R. *Spec. Symp. Am. Soc. Limnol. Oceanogr.* **3**, 555-568 (1980).
18. Morgan, M.D. (ed.) *Ecology of Mysidacea* (Junk, The Hague, 1982).
19. McAndrew, B.J. & Majumdar, K.C. *Zool. J. Linn. Soc.* **80**, 421-435 (1984).
20. Boulenger, G.A. *Catalogue of the Fresh-water Fishes of Africa Vol. 3* (Br. Mus. (Nat. Hist.), London, 1915).
21. Poll, M. *Exploration hydrobiologique du Lac Tanganika 1946-1947. Résultats Scientifiques* Vol. 3, Pt 5A (Bruxelles, 1953).

Does the ocean–atmosphere system have more than one stable mode of operation?

Wallace S. Broecker*, Dorothy M. Peteet† & David Rind†

* Lamont-Doherty Geological Observatory of Columbia University, Palisades, New York 10964, USA

† NASA/Goddard Space Flight Center, Institute for Space Studies, 2880 Broadway, New York, New York 10025, USA

The climate record obtained from two long Greenland ice cores reveals several brief climate oscillations during glacial time. The most recent of these oscillations, also found in continental pollen records, has greatest impact in the area under the meteorological influence of the northern Atlantic, but none in the United States. This suggests that these oscillations are caused by fluctuations in the formation rate of deep water in the northern Atlantic. As the present production of deep water in this area is driven by an excess of evaporation over precipitation and continental runoff, atmospheric water transport may be an important element in climate change. Changes in the production rate of deep water in this sector of the ocean may push the climate system from one quasi-stable mode of operation to another.

MANY feedback loops that may amplify climatic change have been proposed, one of which has recently caught the imagination of palaeoclimatologists. Suggested by the atmospheric CO_2 -content record recovered from the polar ice caps¹⁻³, this scenario involves interactions between climate and ocean circulation⁴⁻⁷.

To appreciate the significance of the observations on polar ice, one must be aware of their context. Over the past several decades studies on deep-sea cores have provided a detailed history of climate during the past million years⁸⁻¹⁰. The first picture that emerged was that continental glaciers have waxed and waned on a timescale of 100 kyr⁹. These climate cycles are asymmetric, involving long periods of glacial buildup suddenly terminated with rapid warmings. Further work revealed that these long intervals of glacial buildup were modulated by cycles of ~20 and ~40 kyr duration^{11,12}, closely allied in timing and relative amplitude to variations in seasonal contrast produced by cyclic changes in the Earth's orbital elements^{11,12}. Orbital forcing is accepted widely as the primary cause of glacial cycles (Fig. 1). Because the exact linkage between changes in seasonal-

ity and changes in climate has yet to be established, however, the issue remains open. Although it is plausible that nonlinearities in the response of snow and ice cover to seasonal variations drive glacial growth and retreat¹³, it is not clear whether this mechanism alone is sufficient to do so.

Ice-core record

Scientists examining the oxygen isotope record preserved in ice cores from Greenland and Antarctica (see Fig. 2), checked whether the major features of this record matched those found in the deep sea. In one sense they did match. Glacial conditions prevailed from before 50 to ~10 kyr BP at which time a transition to interglacial conditions occurred. Once this transition was complete, climate remained remarkably uniform. No evidence for the 20 and 40 kyr cycles characteristic of orbital forcing appears in the ice-core record, in part because the record extends to only ~100 kyr BP and in part because of the exponential foreshortening with depth of the thickness of ice representing a single year. This flow-induced distortion produces a considerable uncertainty in the ice-core chronology before 10 kyr BP.

Although not showing the orbital periodicities, the ice-core record for glacial time does show many brief events during which climatic conditions returned about halfway to their interglacial state. Because such events would be blurred totally by the stirring of worms in all deep-sea sediments with typical accumulation rates (that is, a few centimetres per kyr), this finding is not in conflict with the marine record. However, as

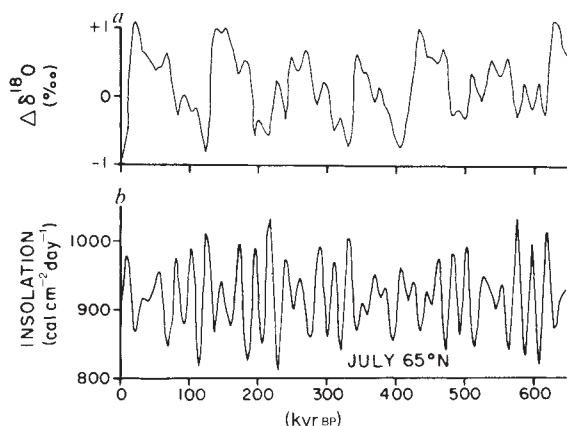
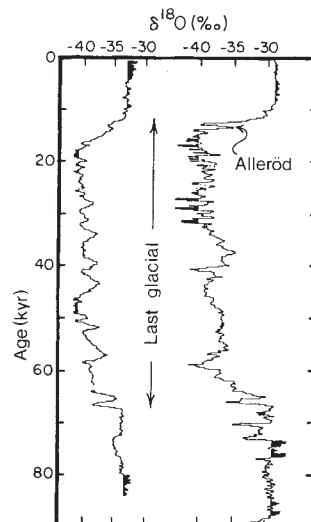


Fig. 1 *a*, Composite of the $^{18}\text{O}/^{16}\text{O}$ record for planktonic foraminifera from several deep-sea cores¹². This record primarily reflects changes in the amount of continental ice in the Northern Hemisphere. The more positive the $\delta^{18}\text{O}$ value, the more ice. *b*, Variation with time of the July insolation at 65°N (ref. 51). The rapid disappearances of the ice that terminate the episodes of very large ice cover correspond to prominent peaks in summer insolation for the latitudes at which the excess ice was located. In addition, there is a strong coherence between the ~20 and ~40 kyr spectral components of ice volume and insolation records. Taken together these ties between insolation and ice volume have convinced most scientists concerned with palaeoclimates that the Earth's glacial cycles are driven by changes in seasonality brought about by cyclic changes in the Earth's orbital elements.

Fig. 2 Oxygen isotope records for ice cores⁴³ from Camp Century, Greenland (on the right) and Byrd Station, Antarctica (on the left). These records reflect mainly changes in air temperature over the ice cap. The more negative the $\delta^{18}\text{O}$ value, the colder the air temperature. The timescales are very approximate as there are as yet no direct radiometric ages for the glacial sections of the ice. Note that the glacial portion of the Greenland record shows a number of events of <1,000 yr duration. These events are muted (or absent) in the Antarctic record.



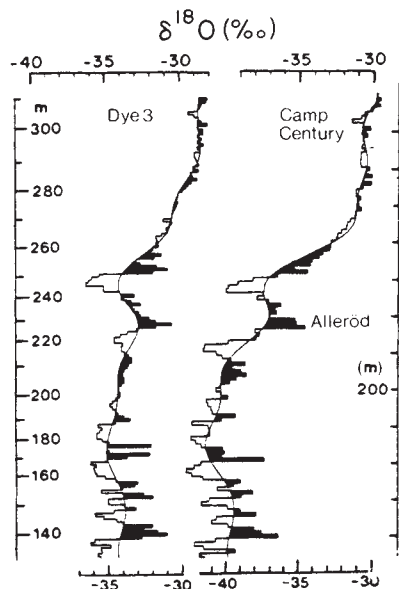


Fig. 3 Comparison between the $^{18}\text{O}/^{16}\text{O}$ records for the Camp Century and Dye 3 sites in Greenland for the time period ~ 40 to ~ 8 kyr BP. For each core a depth scale (metres above bedrock) is given. After some adjustment of the Dye 3 depth scale a good match is achieved between the records in the two cores. (Results obtained by Dansgaard *et al.*¹⁴.)

these idiosyncrasies of the ice-core record were not seen in other records, the initial temptation was to pass them off as climate 'noise' without global significance. A rapid succession of findings has since changed this view of the noise, now the focus of much interest.

First, $^{18}\text{O}/^{16}\text{O}$ measurements by Dansgaard *et al.* on a second long core from Greenland¹⁴ revealed the same events as the first (see Fig. 3), indicating that the events were both real, and occurred throughout Greenland. Second, the events were recorded not only by $^{18}\text{O}/^{16}\text{O}$ ratio changes but also by dust¹⁵, ^{10}Be (ref. 16), SO_4 , NO_3 and Cl content¹⁷ changes in the ice. The ice formed during the warm events is similar to interglacial ice with respect to these properties. Third, the $^{18}\text{O}/^{16}\text{O}$ pattern found for the most recent Greenland event was also recorded in CaCO_3 from lake sediments in Switzerland¹⁸, which gave credence to a correlation made early on by Dansgaard *et al.*¹⁹ (see Fig. 4). In interpreting their $^{18}\text{O}/^{16}\text{O}$ results for the Camp Century Greenland ice core, they suggested that the strong oscillation just before the onset of postglacial time was a record of the Allerød-Younger Dryas oscillation seen in pollen records throughout Europe. During the Allerød, trees abruptly replaced the grasses and shrubs that had characterized glacial time, only to be replaced by the cold flora characteristic of glacial time (see Fig. 4). This so called 'Younger Dryas' cold period lasted ~ 800 yr (that is, from ~ 11 to ~ 10.2 kyr BP) before Holocene forests appeared. However, these results did not influence broader thinking about glacial cycles and their causes.

A few years earlier Berner *et al.* found that the CO_2 content of the air trapped in ice from the glacial sections of the Greenland and Antarctic cores was about two-thirds that for air trapped in ice from the interglacial sections¹. Intrigued by the $^{18}\text{O}/^{16}\text{O}$ events found in the ice and the apparent correlation of the last of these with an oscillation in plant cover seen throughout Europe, Oeschger and co-workers found that there were CO_2 changes associated with these events^{20,21}. In the Dye 3 core, oscillations (as recorded by the $^{18}\text{O}/^{16}\text{O}$ ratio) were accompanied by CO_2 oscillations of ~ 60 p.p.m. (see Fig. 5). The CO_2 change was about two-thirds of the change that accompanied the major glacial-interglacial transition¹⁻³. Although Fig. 5 shows only the CO_2 records for the oscillations within glacial time, a similar CO_2 change is found during the Allerød-Younger Dryas interval at the end of glacial time²⁰.

Oeschger and co-workers are now searching for the CO_2 events in ice cores from the Antarctic. Whereas the other changes found in Greenland (that is, ^{18}O , ^{10}Be , dust, etc.) may have been restricted to the region under the influence of the northern Atlantic, the CO_2 change must, of course, be global. Until confirmed in Antarctic ice, the possibility remains that the sharp changes in CO_2 content found in Greenland are artefacts of summer melting, which may have occurred during the warm events as it did at this low elevation site during much of the Holocene. We emphasize that although it was the discovery of CO_2 events that prompted serious consideration of possible ocean circulation-climate linkages, our speculations below about possible multiple climatic modes do not depend on the validity of the CO_2 changes.

Causes of atmospheric CO_2 change

The finding from ice-core studies that the CO_2 content of air extracted from glacial ice was lower than that extracted from postglacial ice¹⁻³ stimulated investigations as to the origin of natural CO_2 changes, which must be rooted in ocean chemistry^{22,23}. The first scenario for the 90- μatm rise in CO_2 content at the close of glacial time called for the deposition of organic residues on the coastal shelves during the transgression of the sea that accompanied deglaciation. The subsequent finding of brief CO_2 events within glacial time, however, led to a re-examination of the question. If these CO_2 changes are typical of the atmosphere, then explanations involving the transfer of nutrient substances between the ocean and its shelf sediments must be abandoned. Such transfers could certainly not occur fast enough to produce 60-p.p.m. CO_2 changes in ~ 100 yr. To be viable, hypotheses will have to invoke redistributions of C, P and fixed N in the sea. Such redistributions are feasible if the ocean mixing and/or biological cycling change substantially^{4-7,24}. This explains why CO_2 measurements on ice cores

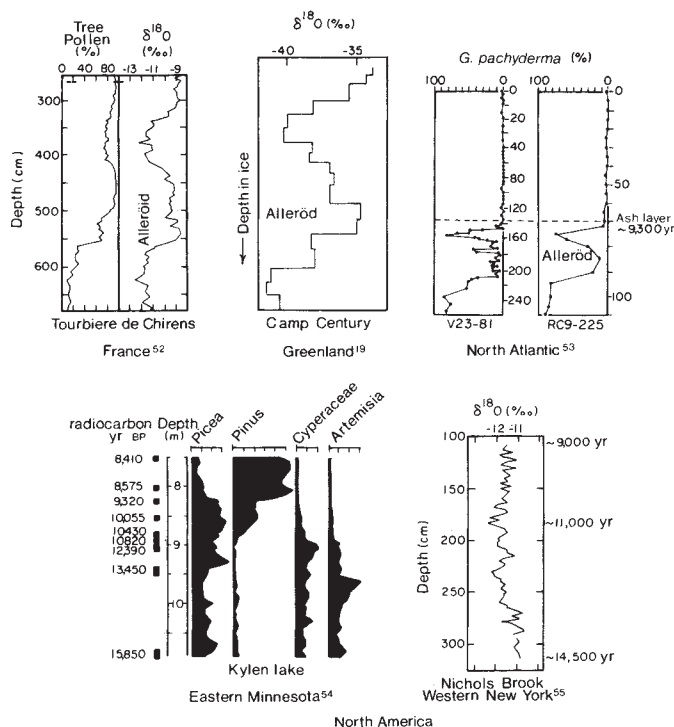


Fig. 4 Examples of records for the time period 13-9 kyr BP from Europe (pollen and $^{18}\text{O}/^{16}\text{O}$ in CaCO_3), Greenland ($^{18}\text{O}/^{16}\text{O}$ in ice) and the North Atlantic (% *G. pachyderma* in the planktonic foraminifera). All show the pronounced Allerød-Younger Dryas oscillation. By contrast no such oscillation is seen in either pollen diagrams or $^{18}\text{O}/^{16}\text{O}$ records from the United States.

suggested links between climate and the joint operation of the atmosphere and ocean system.

Three quite different scenarios have been proposed. Broecker²⁴ pointed out that the evidence from deep-sea cores cited in support of his shelf-deposition hypothesis could also be explained by a change in the C to N and/or C to P ratio in the organic residues falling to the interior of the sea. A 30% decrease in this ratio was needed to explain the CO₂ drop at the end of glacial time.

Broecker and Takahashi⁴ pointed out that a change in the transfer rate of water between the cold- and warm-water spheres of the ocean would produce a redistribution of ΣCO₂ between these reservoirs and a consequent change in atmospheric CO₂ content. The reason is that the biological 'pumping' of carbon by organisms from the surface ocean to the deep ocean is compensated partly by a flow of CO₂ through the atmosphere from the cold- to the warm-water sphere. Hence, a change in the rate of transfer of water between these spheres alters the balance between these two competing effects. To explain the lower CO₂ content during glacial time, Broecker and Takahashi invoked a several-fold higher rate of transfer of water between these spheres.

Three groups⁵⁻⁷ have arrived independently at a different ocean scenario. Having pointed out that a reduction in the NO₃ and PO₄ contents of polar surface waters (that is, more effective biological use) would produce the desired glacial CO₂ reduction, all three groups demonstrated that a decrease in the rate of exchange between polar surface waters and the rest of the ocean may result in reduced nutrient content of polar surface waters and, in turn, of the CO₂ content of the atmosphere.

As the Broecker-Takahashi (B-T) ocean mixing scenario requires enhanced mixing during glacial time whereas the Princeton-Bern-Harvard (P-B-H) scenario requires reduced mixing, the balance of evidence must be carried by marine sediments, which would allow one of these hypotheses to be eliminated. Indeed, the P-B-H hypotheses predict higher than Holocene δ¹³C values for planktonic forams grown in the Antarctic during glacial time whereas the B-T hypothesis predicts lower values. Unpublished ¹³C results by N. Shackleton and R. Fairbanks on planktonic foraminifera from Antarctic deep-sea sediments strongly suggest that the reduction in surface water nutrient content required by the P-B-H hypotheses for glacial time did not occur. Radiocarbon age differences (determined by accelerator mass spectrometry) between planktonic and benthic forams from glacial horizons in deep-sea sediments eventually may provide another basis for making a distinction^{25,26}. It is possible that neither of these scenarios is correct.

Although little is known about either ocean circulation or biological cycling in the past, one important clue has been found.

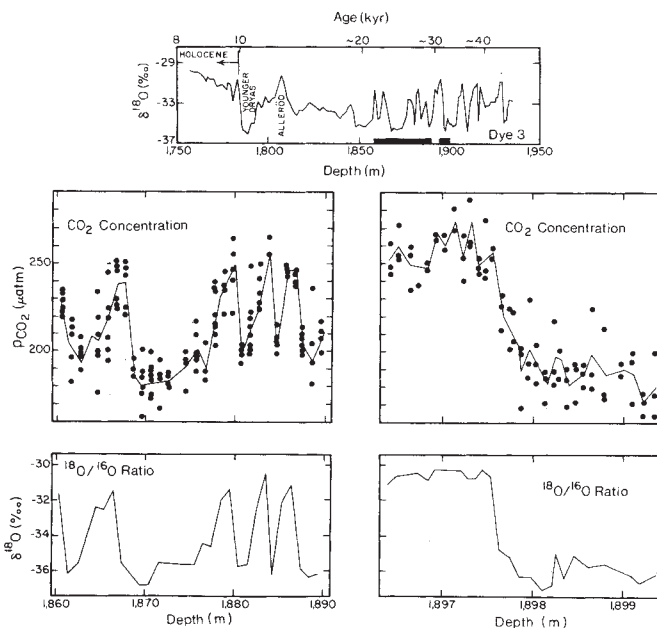


Fig. 5 Detailed PCO₂-δ¹⁸O comparison for two sections of the Dye 3 Greenland core^{20,21}. The location of these sections are shown in the upper panel by the dark bands. The longer band represents a time interval of ~8 kyr and the shorter a time interval of ~2 kyr. The sudden air temperature warmings indicated by the rise in ¹⁸O are accompanied by sharp rises in CO₂ content of air trapped in the ice. (CO₂ results obtained by Oeschger *et al.*)

Evidence from faunal, chemical²⁷ and isotope studies^{23,28-31} of deep-sea sediments suggests that the production of deep water in the northern Atlantic, of great importance to the present-day circulation scheme, was reduced greatly during the peak glacial time. Most convincing is the finding by Boyle that the Cd content of benthic foraminifera from the Atlantic Ocean was higher during glacial time than today²⁷. In the present-day ocean, the Cd and PO₄ contents of various water types in the sea are correlated almost perfectly. Thus, the Cd distribution in the glacial ocean is a good substitute for the P distribution. Boyle has demonstrated that the Cd/Ca ratio in benthic foraminifera is proportional to the corresponding ratio in the water in which they grow. In the present-day ocean, the low PO₄ and Cd content of Atlantic relative to Pacific deep water is attributable directly to the formation of deep water in the northern Atlantic. The reduction in this difference (suggested by the glacial foram results) is thus indicative of a reduction in the magnitude of the deep-water production in the northern Atlantic at that time.

Taken together, these arguments for the reduction of deep-water production in the northern Atlantic during glacial time are convincing. Is it possible then that the brief warm events recorded in the ice cores represent periods during which the glacially weakened northern Atlantic deep-water source was rejuvenated?

Distribution of oscillations

Adequate records are available currently to provide a picture of the geographical distribution for only the most recent fluctuation. This last oscillation appears in the records from both the Camp Century and Dye 3 cores from the Greenland ice cap and is found in pollen and ¹⁸O/¹⁶O ratio records from bog and lake sediments in western Europe. Evidence is also found in sediments in the northern Atlantic with an unusually high accumulation rate. Ruddiman and McIntyre showed that ~13 kyr BP the boundary between polar waters and temperate waters moved from its glacial position to near its present position (see Fig. 6), and remained there ~1 kyr before returning to its glacial position for several hundred years³². Finally, it returned to its present position where it remained throughout the Holocene.

Table 1 Freshwater budget for the North Atlantic Basin, including the Arctic, Mediterranean and Caribbean

Source	Input rate fresh water (m yr ⁻¹)	Freshwater flux (sverdrup)
River runoff ⁵⁰	+0.21	+0.33
Precipitation ⁵⁰	+0.87	+1.36
Evaporation ⁵⁰	-1.21	-1.89
Mediterranean exchange*	-0.07	-0.10
Net†	-0.20	-0.30

The combined area of the North Atlantic basin (0°-80° N) excluding the Mediterranean Sea is 4.9 × 10¹³ m². For freshwater flux, 1 sverdrup represents a flow of a million cubic metres per second.

* The exchange between the more salty waters of the Mediterranean Sea and the less salty waters of the open Atlantic leads to a loss of fresh water from the Atlantic. This process is driven by the high evaporation rate (relative to precipitation and runoff) for the Mediterranean Sea.

† Using a deep water production rate in the northern Atlantic of 20 × 10⁶ m³ s⁻¹ and a salinity excess for this water of 0.50‰, a fresh water flux of 0.3 sverdrups would be obtained, indicating that no large discrepancy exists between oceanographic and meteorological budgets.

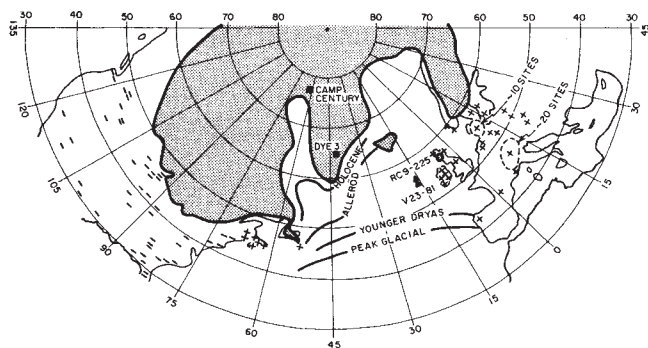


Fig. 6 Map showing the locations of sites at which sediments spanning the 13–9 kyr BP time interval have been studied^{33–36}. At sites designated by a+ an oscillation in climate corresponding to the Alleröd–Younger Dryas is found. At sites designated by a– this oscillation has not been reported. The stippling corresponds to the area covered by ice just before the onset of the Alleröd³⁶. Positions of the polar front in the northern Atlantic during peak glacial, Alleröd, Younger Dryas and Holocene time, as reconstructed by Ruddiman and McIntyre³², and the locations of the two long ice borings in Greenland and of the two ocean cores referred to in Fig. 4 are shown.

There are >50 European sites whose pollen records contain an oscillation in the 13–10 kyr period³³. In contrast to the European pollen record is the lack of evidence for this event in the United States³⁴. Recent research on cores from lakes and bogs in New Brunswick, Nova Scotia and Newfoundland^{35,36}, however, indicate that the oscillation was felt in the eastern regions of maritime Canada. This evidence is summarized in Fig. 6. Thus, as proposed by Mercer³⁷, we may be dealing with a regional change in climate involving primarily the area under the climatic influence of the northern Atlantic Ocean. Because evidence for such an oscillation has been found elsewhere, that is, in the cordillera of South America^{38–41} and in New Zealand⁴², the situation may well be more complicated. This event is muted or absent in the $^{18}\text{O}/^{16}\text{O}$ ratio for the Byrd Station Antarctica record (see Fig. 2)⁴³.

Climate impact of deep-water production

The turning on and off of deep-water production in the northern Atlantic would be expected to produce a regional climate change. The rate of production of deep water is now ~20 Sverdrups (that is, $20,000,000 \text{ m}^3 \text{ s}^{-1}$)⁴⁴. The water feeding the source region for the deep water has a temperature of ~10 °C, whereas the new deep water leaving the region has a temperature of ~2 °C. Thus, as a byproduct of deep-water formation ~ 5×10^{21} cal of heat are released to the atmosphere each year, an amount corresponding to ~30% of the solar heat reaching the surface of the Atlantic Ocean in the region north of 35° N.

To test the geographical distribution of the impact of turning on and off this source of heat, an experiment was performed using the general circulation model for the atmosphere developed by the Goddard Institute for Space Studies. A comparison was made between the air temperature for a model run with present-day ocean conditions and for a model run with the surface ocean temperatures in the region north of the glacial polar front (see Fig. 6) cooled to the levels estimated by the CLIMAP (1981) reconstruction for glacial time⁴⁵. The results of this experiment will be published elsewhere. The air-temperature anomaly generated by this change in sea surface temperature spreads across Europe. A cooling is seen in extreme north-eastern North America whereas no corresponding anomaly is seen in the United States or over Antarctica. The results are, therefore, consistent with the pollen and ice-core evidence. Although this experiment does not prove that the Alleröd–Younger Dryas event was caused by changes in the rate of deep-water production in the northern Atlantic, it does suggest that were such changes to have occurred they

would have produced climate changes with the observed regional pattern.

Deep-water production changes

The most difficult question is why the production of deep water in the northern Atlantic may have resumed for brief intervals during glacial time. For an answer, we must consider the processes important to the present-day production of deep water in the northern Atlantic and for the absence of equivalent production in the northern Pacific. Oceanographers agree that the contrast between the two northern oceans is rooted in the large salinity difference. The waters of the northern Atlantic are saltier than the mean for the upper ocean, whereas the waters of the Pacific are fresher. The Atlantic's extra salt allows waters of density as great as any found in the deep sea to be generated during the winter months. By contrast, even if cooled to their freezing point, northern Pacific waters achieve a density no greater than that of intermediate depth water.

An approximate fresh water budget for the northern Atlantic and its satellite sea, the Mediterranean, is given in Table 1. Despite the fact that this region receives a sizeable portion of the world's river runoff, it loses an even greater amount of water through an excess in evaporation over precipitation. The resulting net loss of fresh water leads to the salt enrichment observed in the northern Atlantic. The salt budget for the North Atlantic must be balanced by the export of salty deep water to the Pacific and Indian Oceans. On a global scale, the water (and salt) budgets of the surface ocean must balance; thus, the fresh water lost from the Atlantic must re-enter the ocean elsewhere. Polewards of 30° N, atmospheric transport produces a net divergence of water vapour over the North Atlantic and convergence over the North Pacific east of the dateline⁴⁶. Thus, if we are to understand what maintains deep water production in the present-day ocean, we must understand why fresh water is being pumped through the atmosphere from the North Atlantic to the North Pacific.

Warren, addressing this question⁴⁷, has shown that this pumping can be explained nicely by the temperature difference between the two oceans. The waters north of 30° latitude in the Atlantic are warmer than their Pacific counterparts. Thus, water is being distilled off the 'warm' Atlantic and condensed on the 'cold' Pacific. If so, then it is the temperature difference on which we must focus. What maintains it?

To some extent the warmer temperatures for the surface of the northern Atlantic must be related to the deep-water formation process. Upper waters in the Atlantic are advected northward into the regions of deep-water production. This advection carries heat to high latitudes. By contrast, in the northern Pacific, deep water upwells to the surface and is carried southward in the upper ocean. In this sense the situation is self-sustaining. Stommel⁴⁸ and Rooth⁴⁹ have discussed the possibility that thermohaline circulation has more than one stable mode. The deep-water circulation pattern maintains surface temperature of the northern Atlantic at a higher value than that for the Pacific. This temperature difference causes the Atlantic to be enriched in salt, whose excess leads to deep-water formation. At steady state, deep-water production must carry away just the amount of excess salt left behind by the atmospheric transport of water from the Atlantic to the Pacific. The standing excess of salt in the northern Atlantic is presumably just high enough to allow the generation of water sufficiently dense to enter the deep Atlantic.

This brings us to the fundamental question of this review—does the ocean–atmosphere system have more than one mode of operation? The idea that the Earth could, under the same solar heating regime, have quite different temperatures depending on its ice albedo or atmospheric greenhouse gas content is not new. Climate modellers have long been aware of the white Earth catastrophe scenario. If somehow the Earth were to be covered with ice, the greatly increased reflectivity of its surface would cause the temperature to fall below the freezing point and thereby stabilize the change. They have also considered the

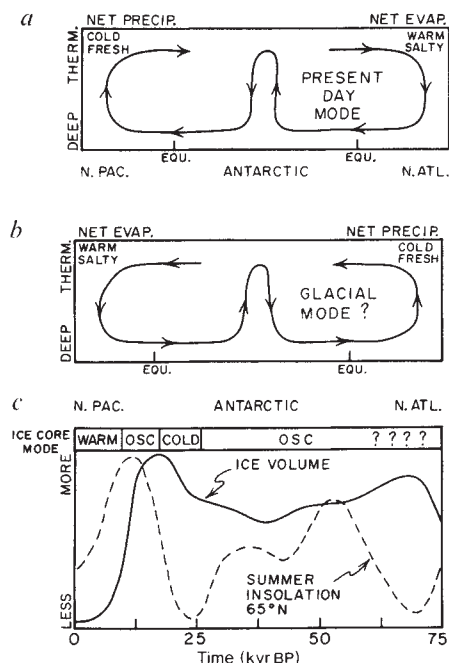


Fig. 7 *a, b*, Two possible extreme modes of ocean circulation are proposed here, corresponding to the warm and cold intervals seen in the ice-core record. *c*, The times when the ice-core record shows the system to have been locked in its warm mode, to have been locked in its cold mode or to have been oscillating between these modes are shown with the ice-volume record and the Northern Hemisphere high-latitude summer insolation record. The intervals of warm Greenland climate correspond to intervals of low ice volume and high summer insolation and the intervals of cold Greenland climate correspond to times of high ice volume and low summer insolation. The intervals when Greenland climate oscillated between warm and cold corresponded to times when the ice volume-summer insolation product had intermediate values. This suggests that the atmospheric transport of water vapour between the Atlantic and Pacific changes with ice cover and summer insolation.

greenhouse catastrophe. If somehow the carbon locked up in sediments were to be released to the atmosphere, the enhanced infrared absorption would raise the planetary surface temperature to the point where carbonate minerals and organic residues could no longer form and thereby stabilize the change. Based on the CO_2 results from ice cores, Oeschger suggested that the Earth had two modes of ocean-atmosphere-biosphere-cryosphere operation. Each oscillation observed in the ice cores involves a jump from one mode of operation to the other.

It is tempting to speculate that one of Oeschger's two modes corresponds to the situation when deep-water production in the northern Atlantic is 'strong' and the other to the situation when deep-water production is 'weak'. If this is the case, then the palaeoclimatic record tells us that, for the summer insolation and ice-cover situation that prevailed during the full interglacial time, the system was in the 'strong' mode and for the corresponding situation that prevailed during full glacial times it was in the 'weak' mode. Because in the present-day ocean the deep-water production corresponds to salt enrichment in the Atlantic, it is tempting to postulate that during full glacial time the opposite was true. Then, salt was being enriched in the Pacific and deep-water flow was the reverse of what it is now (see Fig. 7). Based on oxygen isotope studies on North Pacific cores, deep water may have been produced in the Pacific Ocean during glacial time (N. Shackleton, personal communication). On the other hand, it is possible that the rate of salt enrichment in the Atlantic was considerably less during glacial time, causing a reduction in deep-water production rather than a reversal in flow. Boyle's Cd data favours this latter point of view.

The Oeschger climate oscillations occur in intervals with climates lying between those characterizing full glacial and full interglacial time. If, as the summer insolation at high northern latitudes increased and once again the atmospheric vapour transport balance favoured salt buildup in the Atlantic, then the production of deep water would resume (or intensify), releasing excess heat to the atmosphere over the northern Atlantic. This excess heat would lead eventually to increased discharge of ice from the caps surrounding the Atlantic. If 3% of the excess ice present during full glacial time were added to the Atlantic Ocean by melting during a 100 yr interval, the mean freshwater flux would be 0.6 sverdrups (that is, twice the present imbalance between fresh water and loss for the Atlantic). Thus, the melt water generated in this way would significantly reduce deep-water production, returning northern Atlantic climates to their glacial state. This, in turn, would halt ice retreat and allow the northern Atlantic salt buildup to resume. Thus, it is possible that during periods of ice retreat the system oscillates between the North Atlantic 'strong' and 'weak' modes of deep-water production. Although the physics of such an oscillation has yet to be elucidated, two elements are evident. The first involves water storage in and release from the large Northern Hemisphere glacial ice caps adjacent to the Atlantic. The other involves the transport of salt through the deep sea. As water now resides for many hundreds of years in the deep sea, transients in the deep-ocean salinity distribution created by ventilation changes may be an important element in the oscillator.

Climate modelling

Until now, our thinking about past and future climate changes has been dominated by the assumption that the response to any gradual forcing will be smooth. But if, as proposed by Oeschger, the system has more than one quasi-stable mode of operation, then the situation is more complex. Present general circulation models will at best allow us to study only the changes that will take place if the system remains in its current operational mode. Thus, if the changes that characterized glacial time and those that will characterize the coming CO_2 superinterglacial time involve mode switches, investigations of the transient climate response have to allow for this possibility.

Despite the tenuous nature of the information presently available and of the difficulties inherent in thinking in terms of mode changes, we must begin to explore this alternate track. The programme elements required are clear. At least we must have a joint ocean-atmosphere general circulation model that will allow the exploration of several possible connections between planetary radiation, atmospheric water transport, ocean circulation, atmospheric trace gas content and the marine and terrestrial biosphere. If we are to have confidence that these models provide reliable analogues to the real world, we must gather more information about how each of the important subsystems operates in the present. Finally, we must extract all possible information from the palaeoclimatic record. Even given the full use of present resources, it will be several decades before we possess sufficient understanding to predict future climates. Unless we intensify research in these areas, the major impacts of CO_2 will occur before we are prepared fully to deal with them.

1. Berner, W., Stauffer, B. & Oeschger, H. *Nature* **275**, 53-55 (1979).
2. Delmas, R., Ascencio, J.-M. & Legrang, M. *Nature* **284**, 155-157 (1980).
3. Neftel, A., Oeschger, H., Schwander, J., Stauffer, B. & Zumbunn, R. *Nature* **295**, 220-233 (1982).
4. Broecker, W. & Takahashi, T. *Climate Processes and Climate Sensitivity* (ed. Hansen, J. & Takahashi, T.) 314-326 (*Geophys. Monogr.* **29**, Am. Geophys. U. 1984).
5. Sarmiento, J. & Toggweiler, R. *Nature* **308**, 621-624 (1984).
6. Siegenthaler, U. & Wenk, Th. *Nature* **308**, 624-626 (1984).
7. Knox, F. & McElroy, M. *J. geophys. Res.* **89**, 4629-4637 (1984).
8. Emiliani, C. *J. Geol.* **63**, 538-578 (1955).
9. Broecker, W. & Van Donk, J. *Rev. Geophys. space Sci.* **8**, 169-198 (1970).
10. Shackleton, N. & Opdyke, N. *Quat. Res.* **3**, 39-55 (1973).
11. Hays, J., Imbrie, J. & Shackleton, N. *Science* **194**, 1121-1132 (1981).
12. Imbrie, J. *et al. Milankovitch & Climate I* (eds Berger, A. *et al.*) 269-305 (Reidel, Dordrecht, 1984).
13. Imbrie, J. & Imbrie, J. *Z. Science* **207**, 943-953 (1980).
14. Dansgaard, W. *et al. Science* **218**, 1273-1277 (1982).
15. Dansgaard, W. *et al. Am. Geophys. Un. Monogr. Ser.* **29** (*M. Ewing Symp.* **3**), 288-298 (1984).
16. Beer, J. *et al. Ann. Glaciol.* **5**, 16-17 (1984).
17. Finkel, R. & Langway, C. *Earth planet. Sci. Lett.* (in the press).

18. Siegenthaler, U., Eicher, U., Oeschger, H. & Dansgaard, W. *Ann. Glaciol.* **5**, 149–152 (1984).
19. Dansgaard, W., Johnsen, S., Møller, J. & Langway, C. *Science* **166**, 377–381 (1969).
20. Oeschger, H. *et al. Am. Geophys. Un. Monogr. Ser.* **29** (M. Ewing Symp. 3), 299–306 (1984).
21. Stauffer, B., Hofer, H., Oeschger, H., Schwander, J. & Siegenthaler, U. *Ann. Glaciol.* **5**, 160–164 (1984).
22. Broecker, W. in *Climate Variations and Variability: Facts and Theory* (ed. Berger, A.) 109–120 (Reidel, Dordrecht, 1981).
23. Broecker, W. *Prog. Oceanogr.* **11**, 151–197 (1982).
24. Broecker, W. *Geochim. Acta* **46**, 1689–1705 (1982).
25. Broecker, W., Mix, A., Andree, M. & Oeschger, H. *Nucl. Instrum. Meth. Phys. Res. B5*, 331–339 (1984).
26. Andree, M. *et al. Nucl. Instrum. Meth. Phys. Res. B5*, 340–345 (1984).
27. Boyle, E. & Keigwin, L. *Science* **218**, 784–787 (1982).
28. Duplessy, J., Chenouard, L. & Vila, F. *Science* **188**, 1208–1209 (1975).
29. Kellogg, T., Duplessy, J. & Shackleton, N. *Boreas* **7**, 61–73 (1978).
30. Shackleton, N. *The Fate of Fossil Fuel CO₂* (eds Andersen, N. & Malahoff, A.) 401–427 (Plenum, New York, 1977).
31. Shackleton, N., Imbrie, J. & Hall, M. A. *Earth planet. Sci. Lett.* **65**, 233–244 (1983).
32. Ruddiman, W. F. & McIntyre, A. *Palaeogeogr., Palaeoclimatol., Palaeoecol.* **35**, 145–214 (1981).
33. Watts, W. *Studies in the Late-Glacial of North-west Europe* (eds Lowe, J., Gray, J. & Robinson, J.) 1–21 (Pergamon, Oxford, 1980).
34. Wright, H. (ed.) *Late-Quaternary Environments of the United States Vols 1 and 2* (University of Minnesota Press, 1983).
35. Anderson, T. & Macpherson, J. *6th IPC Conf.* (Calgary, 1984).
36. Mott, J., Grant, D., Stea, R. & Ochiatti, S. *6th IPC Conf.* (Calgary, 1984).
37. Mercer, J. *Arctic Alp. Res.* **6**, 227–236 (1969).
38. Van der Hammen, T., Barends, J., de Jong, H. & De Veer, A. A. *Palaeogeogr., Palaeoclimatol., Palaeoecol.* **32**, 247–340 (1981).
39. Mercer, J. H. & Palacios, O. *Geology* **5**, 600–604 (1977).
40. Wright, H. E. *Quat. Res.* **21**, 275–285 (1984).
41. Heusser, C. J. *Quat. Res.* **22**, 77–90 (1984).
42. Burrows, C. J. *Palaeogeogr., Palaeoclimatol., Palaeoecol.* **27**, 287–347 (1979).
43. Johnsen, S., Dansgaard, W., Clausen, H. & Langway, C. *Nature* **235**, 429–434 (1972).
44. Broecker, W. *J. geophys. Res.* **4**, 3218–3226 (1979).
45. Climap Project Members *Geol. Soc. Am. Map Chart Ser. MC-36* (1981).
46. Peixoto, J. & Oort, A. in *Variations in the Global Water Budget* (eds Street-Perrott, A. *et al.*) 5–65 (Reidel, Dordrecht, 1983).
47. Warren, B. *J. mar. Res.* **41**, 327–347 (1983).
48. Stommel, H. *Tellus* **13**, 224–230 (1961).
49. Rooth, Claes. *Prog. Oceanogr.* **11**, 131–149 (1982).
50. Baumgartner, A. & Reichel, E. *Die Weltwasserbilanz* Munich (1975).
51. Berger, A. *Astr. Astrophys.* **51**, 127–135 (1977).
52. Eicher, U., Siegenthaler, U. & Wegmüller, S. *Quat. Res.* **15**, 160–170 (1981).
53. Ruddiman, W., Sancetta, C. & McIntyre, A. *Phil. Trans. R. Soc. B280*, 119–142 (1977).
54. Birks, H. & Mathewes, R. *New Phytol.* **80**, 455–484 (1978).
55. Eicher, U. & Siegenthaler, U. *Physische Geographie* **1**, 103–110 (1982).
56. Denton, G. & Hughes, T. *The Last Great Ice Sheets* (Wiley, New York, 1981).

ARTICLES

Powerful extragalactic masers

Willem A. Baan

Arecibo Observatory, PO Box 995, Arecibo, Puerto Rico 00613, USA

Two strong extragalactic OH masers have been detected in the galaxies NGC3690 and Mrk 231 which are in the same class as the megamaser source IC4553 (Arp 220). The amplification model for the background continuum proposed for IC4553 can account for most of the powerful OH maser lines and for part of the H₂O emission. Most masing galaxies have sufficient infrared flux to pump the masing regions, except that the H₂O pumping must occur at a much higher conversion efficiency than the OH pumping.

THE extraordinary strength of the few OH and H₂O extragalactic masers^{1–11} suggests that there are emission processes occurring in these sources which differ significantly from those that occur in galactic masers^{12,13}. By far the most luminous OH maser detected to date is in the source IC4553 (refs 4, 14), whose maser emission is superposed exactly on the continuum source in the nucleus of the galaxy^{14,15}. This strongly suggests the presence of amplification of the flux from the nuclear continuum source by a foreground molecular region. The masing region may have a relatively low gain and may not have been observable in the absence of the background source.

This article presents the recent detection of two powerful OH megamasers in the galaxies NGC3690 and Mrk231. Because of their similarity to the emission in IC4553, most of the powerful extragalactic OH maser emission can be interpreted with the continuum amplification model proposed for IC4553. This concept can be extended to the H₂O maser emission as well.

New detections

A search for OH absorption or emission in 110 galaxies with strong continuum sources in June 1984 used the 91-m telescope of the National Radio Astronomy Observatory (NRAO) in Green Bank, West Virginia, which brings the total number of galaxies searched at Arecibo and Green Bank to 240. A cooled gallium arsenide field-effect transistor receiver with a total system temperature of 27 K was used in combination with a 396-channel autocorrelation receiver, which was split in two banks for the two circular polarizations. The sensitivity of the 91-m telescope is 1.1 K Jy⁻¹ at 18 cm. Very gentle third-order baselines were subtracted from the data.

Strong OH emission was detected in two systems, NGC3690 and Mrk231¹⁶. Several new OH absorption features were detected,

among which are the line NGC3079⁵ and a tentative detection in the H₂O maser galaxy NGC4258. The absorption results will be published separately (W.A.B. *et al.*, in preparation). The results on the new OH maser sources are summarized in Table 1 and Figs 1 and 2.

NGC3690. The nuclear activity in the interacting galaxy pair NGC3690/IC694 (Mrk171) is one of the prime examples of starburst activity¹⁷. The pair of Sc galaxies is depicted in plate 299 of Arp's catalogue¹⁸. NGC3690 exhibits the most luminous optical emission lines of any non-Seyfert Markarian galaxy¹⁹ and is also a very strong infrared source^{20–22}. The radio source in NGC3690 has a triple structure and the 10-μm emission and optical line emission are superposed perfectly on this triple¹⁷. Prominent neutral hydrogen absorption towards the radio nucleus of NGC3690 has been seen at the VLA²³. This broad and deep absorption line has a double structure with a separation of 160 km s⁻¹ and can most readily be interpreted in terms of nearly edge-on absorbing disks (in agreement with the optical image). Optical studies²⁴ also indicate a velocity spread across the nuclear region of ~125 km s⁻¹ increasing to the west. NGC3690 is a strong CO emitter; emission of this gas follows the infrared and radio structure of the galaxy (P. Salomon, personal communication; W.A.B. and R. W. Freund, unpublished results). The CO emission also has two peaks at ~130-km s⁻¹ separation.

The OH emission spectrum of NGC3690 is presented in Fig. 1. The 1,667- and 1,665-MHz lines are clearly visible in the spectrum and are merged slightly. The features have been verified recently using the Very Large Array (VLA)-A. The double structure in the 1,667-MHz line agrees quite well with the two H I absorption features. Assuming the high velocity OH component originates at the west continuum component and

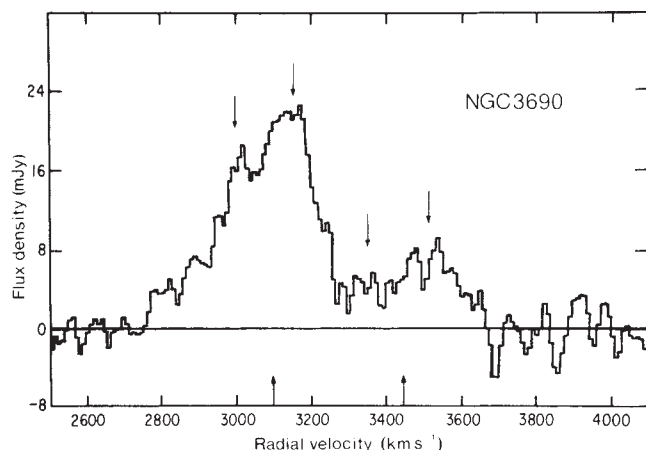


Fig. 1 The OH spectrum of NGC3690 obtained using the 91-m telescope of NRAO. The integration time is 59 min; the velocity scale refers to the 1,667-MHz feature. The 1,665-MHz feature is offset 351 km s⁻¹ from the 1,667-MHz line. The pair of long arrows indicate the centre velocity of the features. The two pairs of short arrows indicate the double structure found in the H I and CO features.

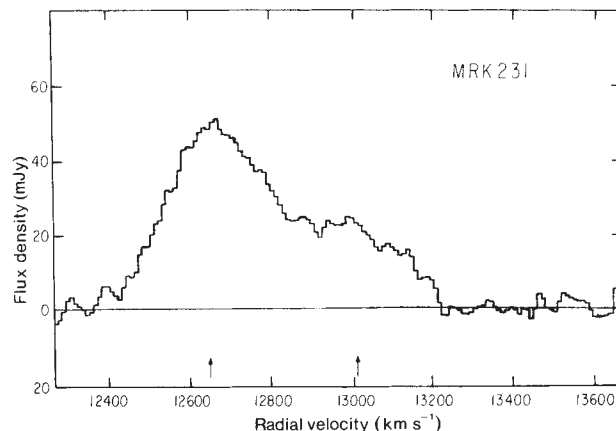


Fig. 2 The OH spectrum of Mrk231. The integration time for the spectrum is 29 min. The arrows indicate the centre velocities for the two features obtained with Gaussian fitting. The H I absorption velocity lies 14 km s⁻¹ below this. The velocity scale refers to the 1,677-MHz feature.

the low velocity component at the east, the optical depths for the two components are ~ -0.08 and -0.15 , using the respective radio continuum fluxes. This suggests a correlation between the infrared flux and the maser gains, because the infrared flux from the west is ~ 2.5 times that from the east.

Mrk231. This is one of the most prominent Sy1 galaxies and is considered to be a prime example of Seyfert properties. However, Mrk231 is unique among Sy1 galaxies for its red colours and its strong interstellar absorption lines²⁵. Most of its luminosity lies in the infrared^{20,21}. The nucleus of Mrk231, having a strong unresolved radio source coincident with the optical continuum peak, cannot be modelled simply with bursts of star formation²⁶. Nevertheless, Mrk231 is one of the few nuclei that shows both Seyfert 1 properties and evidence of star bursts²⁴.

Mrk231 has optical emission-line velocities that are different from the absorption-line velocities²⁷. Broad H I absorption has been found at the lower redshift^{23,28}, suggesting the presence of an edge-on disk structure. Gaussian fits of the OH emission lines detected in Mrk231 (Fig. 2) indicate a velocity which is 14 km s⁻¹ higher than the H I absorption velocity. For an isotropic emission pattern, the emission in the 1,667-MHz line alone is 2,500 $L_{\odot}$, which is 3.5 times greater than that of IC4553. The luminosities quoted for the masers are isotropic. It is, however, probable that the maser emission is not isotropic and that the beaming is different for the OH and the H₂O masers. Assuming that the maser emission in Mrk231 originates in a cloud that exponentially amplifies the continuum flux²⁹, the peak optical depth required for explaining the 1,667-MHz lines is -0.3 . If all emission in Mrk231 occurs in front of the $<0.3''$

compact source the brightness temperature of the 1,667-MHz peak is $>3.6 \times 10^5$ K.

OH megamaser IC4553

The case of the OH maser source in IC4553 may be taken as a basis for the discussion of extragalactic masers. IC4553 (Arp220) has been studied extensively in different wavelength regimes and the data reveal a rather complete picture of the properties of the different masing regions.

The galaxy IC4553 has been detected in emission at all four 18-cm transitions of OH. The two main lines have been presented earlier⁴. The 1,667-MHz feature has a peak strength, which is 0.82 of that of the continuum source, whereas the peaks of the 1,665-, 1,612- and 1,720-MHz lines are respectively 0.24, 0.08 and 0.03 of the 1,667-MHz peak flux (W.A.B., in preparation). VLA-A¹⁴ and MERLIN¹⁵ observations at 18 cm show that the line emission is exactly superposed on a triple continuum structure of $0.6 \times 2.0''$.

The high-resolution results indicate that the bulk of the masing material is part of a rotating molecular disk centred on the nuclear continuum source^{14,15}. A second distinct emission region exists on the eastern side of the source, which has a radial velocity ~ 300 km s⁻¹ higher than the systemic velocity of the molecular disk¹⁴. In H I the galaxy IC4553 displays a very strong and broad (740 km s⁻¹) absorption line³⁰. VLA-A observations³¹ at 21 cm reveal that the bulk of the absorption also results from a rotating disk, which belongs to the same large-scale structure as the molecular disk and is centred on the nuclear continuum radio source. The high-velocity OH emission feature has also a counterpart in H I absorption.

Table 1 OH megamasers

Name	Alternative names	$V_{\text{abs}}(\text{H I})$ (km s ⁻¹)	V_{OH} (km s ⁻¹)	Isotropic luminosity $L_{\text{OH}}(1,667)$ ($L_{\odot}$)	Continuum flux (mJy)	Line flux 1,667 MHz (mJy)	1,665 MHz (mJy)	Hyperfine ratio R_{H}	Optical depth $\tau(1,667)$	Brightness temperature T_{b} (K)	Comments
NGC3690	Arp299	3,115	3,101	85	220	18	4	4.5	-0.08	$\sim 5 \times 10^2$	East component
	Mrk171	(23)			140	21	7	3.0	-0.14	$\sim 5 \times 10^2$	West component
IC4553	Arp220	5,350	5,373	700	278	280	36	5.4	-0.7	3×10^6	Region 1(14)
		5,659 (30, 31)	5,642 (4)			20	8	2.4			Region 2
Mrk231	U08058	12,638 (23, 28)	12,650	2,500	160 (29)	51	22	2.3	-0.28	$>3 \times 10^4$	

References are given in parentheses.

Table 2 Comparison of extragalactic OH and H₂O masers

Name	Alternative name	Distance§ (Mpc)	Nuclear spectral classification	Absorption column density H I N_{H}/T_s	OH $N_{\text{OH}}/T_{\text{ex}}$	Infrared luminosity ($L_{\odot}$)	Luminosity ratio R_c $L_{(1R)}/L_{(B)}$	Isotropic OH line luminosity 1,667 MHz ($L_{\odot}$)	Isotropic H ₂ O line luminosity ($L_{\odot}$)	Radio continuum	
NGC253		3.1	N (24, 36)	3.4×10^{19} (63, 64)	4.7×10^{15} (11)	1.5×10^{10} (20, 21)	—	2×10^{-1} (11)	—	Extended, compact core (56, 63, 64)	Main OH lines
NGC520		42	H (24, 36)	$>2.0 \times 10^{19}$ (3, 30)	4.3×10^{15} (3)	8.6×10^{10} (49)	6	5×10^{-5} (3)	—	Compact, with extension (56)	1,667-MHz OH line
NGC3690	Arp299	56	H (24, 36)	1.1×10^{20} (23)	—	7.4×10^{11} (17, 22)	10 (22)	85 (this paper)	—	Triple compact, extended (17)	Main OH lines
IC4553	Mrk171 Arp220	98	S or H (24, 36)	1.1×10^{20} (30, 31)	—	2×10^{12} (34)	80 (34)	700 (4, 14)	—	Compact triple	Main OH lines
Mrk231	U08058	226	S (24, 36)	2.8×10^{19} (23, 28)	—	4×10^{12} (20, 21, 25)	27 (34)	2,500 (this paper)	—	Compact, weak extension (20)	Main OH lines
M82	NGC3034	3.5	H (24, 36)	2×10^{20} (1, 2)	8×10^{14} (1, 2)	4×10^{10} (20, 21)	8 (34)	3.6×10^{-2} (1, 2)	1.5 (6)	Compact source disk extension (1, 2, 57)	Main OH lines
Circinus		4	S?	—	—	1.2×10^{10} (52, 53)	—	—	35 (10)		
NGC4258		9.5	S, L (24, 36)	—	5×10^{14} (†)	6×10^8 (54)	>0.3	—	157 (5, 6, 7)	Extended S-shaped at 20 cm (58, 59)	
NGC4945		8.1	S(?)	6.6×10^{19} (50)	1×10^{16} (50, 55)	$>2 \times 10^{10}$ (52)	—	3×10^{-5} (55)	240 (7, 8, 9)	Extended at 20 cm	1,612-MHz OH line
NGC1068		20	S	8×10^{17} (51)	$<8 \times 10^{14}$ (39)	3×10^{11} (20)	5	—	320 (6, 7)	Compact S-shaped (56, 60)	
NGC3079		21	L, S (24, 36)	*	7.6×10^{15} (5)	1×10^{11} (53)	4	†	960 (5, 7)	Compact core, extended structure at 20 and 6 cm (61)	1,667-MHz OH line

References are given in parentheses. The marginal detections of H₂O in NGC6946^{6,7} and NGC6240⁷ have not been included in the table.

* H I absorption in NGC3079 may be evident from the published spectrum of Shostak⁶².

† The strength of the OH maser lines in NGC3079 is not yet known.

‡ A tentative detection has been made of OH absorption in NGC4258 (W.A.B., A. D. Haschick & J.T. Schmelz, in preparation).

§ $H_0 = 55 \text{ km s}^{-1} \text{ Mpc}^{-1}$.

The optical appearance of the galaxy IC4553 suggests that the galaxy is definitely the result of a recent merger (see plate no. 220 in Arp¹⁸). Recent optical studies indicate that the dark band, which divides the image in half, is a dust lane^{32,33}. The disturbed appearance of the dust lane suggests a disk structure, similar to that seen in M82 (ref. 33). The outer optical structure resembles that of the merging pair in the 'Antennae' (NGC4038/9). The notion of tidal interactions is supported by the high velocity H I gas (in absorption) and OH gas (in emission). No H I emission has been seen in IC4553 (ref. 31).

IC4553 is a very prominent infrared source as has been found recently by IRAS³⁴ and ground-based instruments³⁵. The infrared source in IC4553, with a minimum size of $\sim 1.4''$ and a spectral temperature of 61 K (refs 34, 35), is centred on the position of the continuum source^{32,34} in the centre of the dust lane. IC4553 is unique among the infrared galaxies because it has the highest known ratio of infrared luminosity to blue luminosity ($R_{\text{ib}} = 80$). This high ratio could indicate that a powerful burst of star formation is in progress. Optical emission lines suggest Seyfert-like (Sy2) properties, but they would not be inconsistent with an H II region spectrum^{24,32,36}. Starburst models as well as Seyfert models (as needed for Mrk231) present difficulties in explaining IC4553 because of the heavy dust obscuration^{24,36}. IC4553 exhibits strong H₂ emission^{37,38}, which could be explained by shock excitation of interstellar clouds. No flux was detected in a 455-min exposure in the short ultraviolet wavelength regime with the IUE telescope (W.A.B. and R. Panek, unpublished results).

There is still controversy as to the nature of the galaxy IC4553. The nucleus may be the advanced stage of a merger of two nuclei or the nuclear activity has been triggered by a tidal interaction, which also resulted in a disturbed outer structure. The radio and infrared sources could be evidence either of Seyfert activity or of a starburst embedded in the very dusty nuclear region.

Working model

On the basis of the observational results on the OH megamaser in IC4553, a model^{14,15} has been suggested in which the bulk

of the maser emission is caused by amplification of the radio flux from the nuclear continuum source by foreground molecular material. Here I apply this model to the two newly discovered OH masers and to extragalactic H₂O masers. The OH and H₂O molecules in molecular clouds may be inverted partially by the infrared radiation field or other pumping processes. This relatively weak population inversion occurs in relatively extended molecular structures, in contrast to the stronger but more localized population inversions around condensed H II regions which characterize masers in our Galaxy. Because of its low exponential gain, the molecular material is not observable as a maser when observed on its own. However, when a background source is placed behind the cloud, it can amplify the continuum flux to an observable level.

To obtain the correct configuration for continuum amplification the following ingredients are required: (1) the presence of a background continuum source, possibly in the centre of the galaxy; (2) a nearly edge-on molecular disk, thus increasing the probability of seeing a molecular cloud with an inverted population along the line-of-sight; (3) a pump for inverting the molecular material.

Below, I address these requirements in more detail and present the current evidence in support of such a model. Some of these characteristics have already been mentioned in relation to the masers in IC4553 (refs 14, 15, 32) and NGC3079 (ref. 5). The main characteristics of all prominent masing galaxies are given in Table 2.

Continuum sources. Most known masing galaxies are characterized by compact nuclear continuum sources, some of which also have an extended nuclear structure. However, the crucial question is where the maser emission occurs relative to the continuum source. In the case of IC4553, the maser emission is exactly superposed on the continuum structure and the emission and continuum contours are unmistakably the same^{14,15}. Spectral line studies of the new OH maser sources at the VLA are not yet available, but the available evidence for NGC3690 strongly suggests that the OH emission originates at the location of the three continuum sources. Mrk231 has a compact nuclear radio source. In the case of M82 the OH maser emission does not

originate at the peak of the continuum map²; but the continuum flux is still considerable at the maser locations and maser amplification is certainly plausible. For the H₂O masers, a coincidence of maser and continuum emission has been found in three cases. The prominent maser feature in NGC3079 (refs 4, 6) coincides with the nuclear source at 2" resolution with the VLA-D (A. D. Haschick and W.A.B., unpublished results). Similarly, the prominent feature in NGC4258 falls on a weak source in the extended continuum structure, whereas the high-velocity feature in NGC1068 coincides with one of the eastern continuum knots (M. J. Claussen, personal communication). VLBI and VLA-A studies are in progress to determine these relative positions.

Molecular disks. The main observable characteristics (besides the optical image) of seeing a galactic disk edge-on are high absorption column densities. Most masing galaxies show evidence of H I and/or OH absorption and some have the highest known H I absorption column densities. The masing galaxy with a low H I absorption column density, NGC1068, also has a relatively low inclination angle; a severe upper limit has been found for its OH absorption column density (ref. 39; W.A.B. *et al.*, in preparation). The molecular disk in our Galaxy is inclined by an angle of 25° relative to the H I disk⁴⁰; this might be true for other galaxies. A recent compilation of all known OH absorption lines indicates a correlation of the OH absorption optical depth and the estimated inclination angle of the galaxy³. This result supports the notion that most of the absorbing OH molecular clouds are confined to a relatively compact disk, approximately aligned with the large-scale disk of the galaxy.

Pumping mechanisms. Here, I assume that the masing molecular gas is contained in molecular clouds that are not necessarily next to compact H II regions. Diffuse molecular gas can have maser action, as has been seen for the galactic molecular clouds along the line-of-sight to the continuum source 3C123 (ref. 41). The 1,720-MHz emission line is found to be stronger on-source than off-source, which can be readily explained with weak amplification at 1,720 MHz of the continuum flux of 3C123 as it passes through the molecular cloud. This is the only known definite example (besides IC4553) of amplification of a background continuum.

The pumping agent for the OH in molecular clouds is probably the far infrared radiation from dust^{42,43}. For IC4553, all four 18-cm OH transitions have been seen in emission, which is unusual because generally either 1,720 MHz or the other three lines are in emission, but not all four. The pumping of H₂O masers is still subject to study. For H₂O/H II masers it is difficult to explain how such large maser fluxes emerge from clouds of such small sizes. Collisional processes may be the dominant pumping mechanisms for H₂O masers¹³.

Most galactic OH and H₂O masers are thought to be saturated^{12,13}, which means that there is a steady production of photons throughout the masing region. Saturated masers have a high efficiency of converting pump photons into maser photons. Unsaturated masers have an exponential gain but have a lower conversion efficiency. However, an unsaturated maser can amplify background continuum radiation. As the line radiation field in the maser gets stronger, the maser will eventually saturate. The presence of a background continuum may bring this about earlier. An observer will then see the (saturated or unsaturated) masers superposed on the continuum source. Without a background, source the masers may not saturate at all and even be unobservable. This situation may exist in the H₂O sources NGC3079 (ref. 5) and NGC4258, where narrow features could represent (mostly) saturated masing pockets of gas, whereas broader features represent (mostly) unsaturated amplification of the radio continuum. The OH maser lines are generally broad and could mainly result from amplification of the background source. Naturally broad OH lines may also have some high brightness saturated features embedded in them. Using the presently known sizes (upper limits) for the continuum sources, one finds a brightness temperature $T_b < 10^7$ K for the masers, which is considerably below those of saturated galactic masers ($T_b = 10^{10}$ – 10^{11} K; ref. 12). Much will be learned from

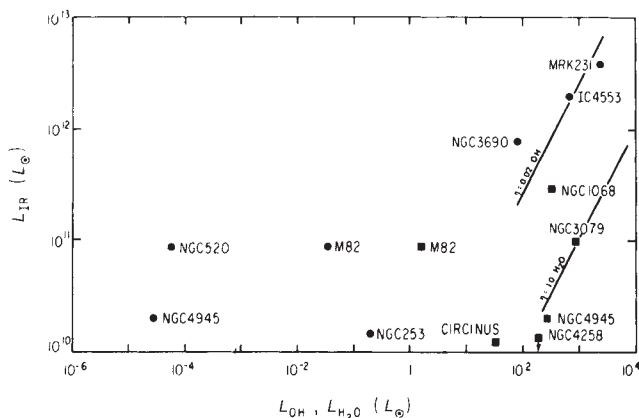


Fig. 3 Relation between the isotropic OH and H₂O luminosities and the integrated infrared luminosity. The circular entries are for OH masers and the square entries are for H₂O masers. The linear relationships in the diagram represent constant conversion efficiency curves for the OH and H₂O masers. For determining these curves the linewidth has been taken to be 300 km s⁻¹ for the OH masers and 100 km s⁻¹ for the H₂O masers.

VLBI observations as they will reveal the compact and more saturated components.

A general problem for galactic masers is the absence of an identifiable pump source, in particular of a sufficient infrared flux. However, this is not the case for the extragalactic maser sources, which tend to have prominent infrared sources mostly at the location of the radio continuum sources (that is, NGC3079, NGC3690, IC4553, Mrk231). In Fig. 3 the OH and H₂O maser luminosities have been plotted against the infrared luminosity of the galaxy. The low luminosity masers seem to be independent of $L(IR)$. However, the high luminosity masers show an approximately linear increase in luminosity with increasing $L(IR)$. Such a linear relationship is expected and observed for saturated stellar masers, but seems also plausible for unsaturated masers. Two pump efficiency curves based on the spectra of the prominent maser galaxies (IRAS point-source catalogue) indicate that the OH masers may have an efficiency of a few per cent for converting photons at 30 μ m into OH maser line photons. However, if the inferred radiation field at 10 μ m pumps the H₂O maser, this should happen at a conversion efficiency of 50–100%. These conversion efficiencies are typical for unsaturated and saturated masers, respectively. Naturally collisional processes can also contribute to the pumping of the water vapour. Evidently, OH and H₂O masers have different pumping mechanisms and efficiencies, or the H₂O masers may be more directional (thus resulting in an overestimate of the luminosity). The low luminosity maser sources may result from similar large-scale processes, but they can also be just like the powerful masers seen in our Galaxy.

OH and H₂O masers

It is apparent from the above discussion that the properties of the OH maser sources and those of H₂O are similar in many ways. Because of these similarities, I suggest that the basic model, amplification of the background source flux, may apply to both types of sources. However, some maser characteristics are different, indicating that different physical conditions prevail in the masing OH and H₂O regions.

In compact H II regions, H₂O masing occurs at higher densities than OH masing (see ref. 12). The H₂O masers are necessarily located closer to the exciting source. If these excitation conditions apply to the situation described here of foreground molecular clouds, then H₂O inversion occurs only in the cores of clouds, whereas OH can be inverted over much large regions. This suggests that: (1) the linewidths of the H₂O maser lines be much smaller than those of the OH maser lines; (2) the H₂O masing regions are likely to be much smaller than

the OH masing regions, and are expected to vary on shorter timescales.

The outstanding H₂O maser features in the spectra of N3079 and N4258 do indeed vary on timescales of months (refs 5, 7; M. J. Claussen, personal communication). The spectra of most H₂O masing galaxies show distinct narrow features, sometimes superposed on some broader feature. Only the weakest OH masers have spectra breaking up into distinct features. However, the prominent narrow H₂O features have linewidths which are significantly larger than those of galactic masers.

Another difference between OH and H₂O maser lines is their velocity relative to the systemic velocity of the galaxy. In IC4553 the H I disk in absorption and the molecular disk in OH emission have the same centre velocity. This is also true for Mrk231 and NGC3690. However, the H₂O masers tend to have velocities away from the centroid velocity of the galaxies. This is strikingly apparent from the comparison of the OH absorption spectrum of NGC3079 and its H₂O spectrum⁷. The H₂O features in NGC3079 are blueshifted by 140 km s⁻¹ relative to the OH centre velocity, but the features coincide in velocity with weak OH emission lines, which are superposed on the 1,667-MHz absorption line (B. E. Turner, personal communication). Similarly, the lines of NGC4258 are blueshifted 60 km s⁻¹ when compared with the tentative OH absorption detection. NGC4945 has redshifted H₂O lines relative to the OH absorption velocity but these lines also coincide in velocity with a 1,612-MHz OH emission feature. M82 has one blueshifted H₂O feature, which again coincides in velocity with a 1,667-MHz OH emission feature, whereas another feature lies far to the red side of the OH absorption.

The prominent H₂O emission features originate possibly in dense (shocked) pockets of molecular gas with relatively high gains and having high relative velocities. In contrast, the OH emission may originate in more extended regions. Most of the galaxies in Table 2 have been searched for both OH and H₂O, but only in M82, NGC4945 and NGC3079 has emission been seen at both frequencies and at identical velocities.

Properties of masing galaxies

Two of the prime characteristics of the sample of masing galaxies, the presence of a strong nuclear radio continuum source and the infrared flux, are likely to be related⁴⁴. Peculiar and interacting spirals have relatively high continuum-flux densities as well as extended nuclear sources^{24,36}. Most of this radio emission can be explained in terms of starburst models⁴⁵. However, it is not clear whether these nuclei have a Seyfert or a starburst nature, since a low ionization or Seyfert spectrum can hide an underlying H II region spectrum completely. The dust can effectively hide from sight the H II regions associated with the starbursts, which in turn may be responsible for the radio continuum sources. Possibly starburst nuclei and Seyfert nuclei are different stages of evolution of objects of one class⁴⁶.

The infrared properties of galaxies are still being elucidated, but it seems that double, interacting or merging galaxies are prominent among infrared galaxies⁴⁷⁻⁴⁹. Of the sample of masing galaxies, most are strong infrared sources and a few are known to be interacting or merging systems (NGC3690, IC4553, NGC520). It is likely, however, that most of these galaxies have experienced recent bursts of star formation and that the origin of the megamaser phenomenon can be traced to the coexistence of a radio continuum source, an ultraviolet radiation field to be converted into an infrared radiation field by the dust and remnant molecular material.

The detection rates of extragalactic masers have been very low. Four of the OH masers have been detected during surveys of 240 predominantly late-type spiral galaxies with radio continuum (W.A.B. *et al.*, in preparation). Recent searches for H₂O emission in 65 spirals produced four new cases⁵⁻⁷. These detection rates of 2-6% have been affected by: (1) a low filling factor for the proper type of molecular clouds (this factor is likely to be larger for the OH gas than for the denser H₂O cloudlets); (2) a relatively short lifetime of masing clouds compared with

the lifetime of the continuum and infrared sources; (3) the relative detectability of signals at OH and H₂O; (4) the composition of the sample. The slightly higher probability of detecting H₂O masers could be caused by tighter selection criteria for the sample or by the fact that the gains for some H₂O maser features are higher, leading to better detectable features.

Conclusions

Maser activity in other galaxies is a newly discovered phenomenon and only a small number of cases have been found. Some of the characteristics of the OH and H₂O megamasers can be explained with a model involving unsaturated maser amplification of radio continuum structures (compact or extended) by foreground molecular material. The highest probability for a galaxy to produce observable maser emission is to have a nearly edge-on molecular disk with an embedded radio source and a proper pumping mechanism for the molecules. Most masing galaxies fulfill these requirements.

The nuclear spectra of the prominent masing galaxies have been classified as H II region or Seyfert types. A scenario for masing sources would then involve some sort of nuclear activity (with a burst of star formation or a Seyfert engine), which provides the radio source and an ultraviolet flux. The ultraviolet flux is downgraded to infrared by the surrounding dust. The infrared radiation is used to pump the foreground molecular gas, which in turn gives the maser emission. The starburst activity in several galaxies has probably been triggered by a merger or a galaxy/galaxy interaction in the recent past. All these nuclei could belong to one class, but have different amounts of dust and are at different stages of evolution.

The properties of the OH and H₂O maser lines are sufficiently different to suggest a difference in the excitation of the two molecules. Although they both may amplify the background continuum, the OH masing regions are probably more extended, whereas the prominent H₂O maser features are likely to originate in smaller cloudlets, either falling into the nucleus or flying out. These smaller pockets have a relatively higher gain (more saturated) than the surrounding (unsaturated) material, which results in an H₂O spectrum with narrow lines superposed on broader features.

The observed infrared flux for the masing galaxies seems to be sufficient to pump the molecular gas. However, for the OH masing galaxies, the infrared photon-to-line photon conversion efficiency can be only a few per cent, whereas for the H₂O masers this must be 50-100%. Possibly the bulk of the H₂O maser emission is due to saturated components. An approximately linear relationship between maser and infrared luminosity can be seen for the OH and H₂O masers. Apparently each type runs at its own approximately constant efficiency. It is of interest to speculate what this diagram tells us about other galaxies to be detected in OH or H₂O. At present H₂O features coincide with weak OH features in three galaxies (M82, NGC3079, and NGC4935). Except for these galaxies, OH and H₂O emission seem to be mutually exclusive.

The similarities and differences between OH and H₂O masers are important for determining the physical conditions of the masing material and provide valuable insight into the nature of these active galactic nuclei. Extragalactic megamasers seem to represent a new form of astronomical 'image processing'. Unlike gravitational lensing, the processing is provided by diffuse clouds of interstellar molecules.

I thank Aubrey Haschick for discussions, Joan Schmelz for assistance with the observations, and Mike Davis for critical reading of the manuscript. The Arecibo Observatory is part of the National Astronomy and Ionosphere Center, operated by Cornell University under contract with the NSF. The National Radio Astronomy Observatory is operated by Associated Universities, Inc., under contract with the National Science Foundation.

Received 29 October 1984; accepted 11 March 1985.

1. Nguyen-Q-Rieu, Mebold, U., Winnberg, A., Guibert, J. & Booth, R. *Astr. Astrophys.* **52**, 467-469 (1976).

2. Weliachew, L., Fomalont, E. B. & Greisen, E. W. *Astr. Astrophys.* **137**, 335-342 (1984).
3. Baan, W. A., Haschick, A. D., Buckley, D. & Schmelz, J. *Astrophys. J.* (in the press).
4. Baan, W. A., Wood, P. A. D. & Haschick, A. D. *Astrophys. J. Lett.* **260**, L49-L52 (1982).
5. Haschick, A. D. & Baan, W. A. *Nature* **314**, 144-146 (1985).
6. Clausen, M. J., Heiligman, G. M. & Lo, K. Y. *Nature* **310**, 298-300 (1984).
7. Henkel, C. *et al. Astr. Astrophys.* **141**, L1-L3 (1984).
8. dos Santos, P. M. & Lepine, J. R. D. *Nature* **278**, 34 (1979).
9. Batchelor, R. A., Jauncey, D. L. & Whiteoak, J. B. *Mon. Not. R. astr. Soc.* **200**, 19p-21p (1982).
10. Gardner, F. F. & Whiteoak, J. B. *Mon. Not. R. astr. Soc.* **201**, 13p-15p (1982).
11. Gardner, F. F. & Whiteoak, J. B. *Mon. Not. R. astr. Soc.* **173**, 77p-81p (1975).
12. Reid, M. J. & Moran, J. M. A. *Rev. Astr. Astrophys.* **19**, 231-276 (1981).
13. Downes, D. *Birth and Infancy of Stars* (eds Lucas, R. & Omont, A.) (in preparation).
14. Baan, W. A. & Haschick, A. D. *Astrophys. J.* **279**, 541-549 (1984).
15. Norris, R. P., Baan, W. A., Haschick, A. D., Booth, R. S. & Diamond, P. J. *Mon. Not. R. astr. Soc.* (in the press).
16. Baan, W. A., Haschick, A. D. & Schmelz, J. T. *IAU Circ. No. 3993* (1984).
17. Gehrz, R. D., Sramek, R. A. & Weedman, D. W. *Astrophys. J.* **267**, 551-562 (1983).
18. Arp, H. C. *Atlas of Peculiar Galaxies* (California Inst. Techn., Pasadena, 1966).
19. Weedman, D. W. *Astrophys. J.* **171**, 5-12 (1982).
20. Rieke, G. H. & Low, F. J. *Astrophys. J. Lett.* **200**, L67-L69 (1972).
21. Allen, D. H. *Astrophys. J.* **207**, 367-375 (1976).
22. Lonsdale, C. J., Neugebauer, G. & Soifer, B. T. *B.A.A.S.* **16**, 470 (1984).
23. Dickey, J. M. *Astrophys. J.* **263**, 87-93 (1982).
24. Heckman, T. M., van Breugel, W., Miley, G. K. & Butcher, H. R. *Astr. J.* **88**, 1077-1087 (1983).
25. Lacy, J. H. *et al. Astrophys. J.* **256**, 75-82 (1982).
26. Ulvestad, J. S. *Astrophys. J.* **259**, 96-102 (1982).
27. Adams, T. F. *Astrophys. J. Lett.* **176**, L1-L3 (1972).
28. Heckman, T. M., Balick, B. & Sullivan, W. T. *Astrophys. J.* **224**, 745-760 (1978).
29. Ulvestad, J. S., Wilson, A. & Sramek, R. A. *Astrophys. J.* **247**, 419-442 (1981).
30. Mirabel, I. F. *Astrophys. J.* **260**, 75-80 (1982).
31. Baan, W. A., van Gorkom, J., Schmelz, J. T. & Mirabel, I. F. *Astrophys. J.* (submitted).
32. Norris, R. P. *Mon. Not. R. astr. Soc.* (submitted).
33. Schild, R. E. *Sky and Telescope* January, 24-45 (1985).
34. Soifer, B. T. *et al. Astrophys. J. Lett.* **283**, L1-L4 (1984).
35. Emerson, J. P. *et al. Nature* **311**, 237-238 (1984).
36. Keel, W. *Astrophys. J.* **282**, 75-84 (1984).
37. Joseph, R. D., Wright, G. S. & Wade, R. A. *Nature* **311**, 132-133 (1984).
38. Becklin, E. *et al. IAU Circ. No. 3968* (1984).
39. Rickard, L. J., Bania, T. M. & Turner, B. E. *Astrophys. J.* **252**, 147-155 (1982).
40. Liszt, H. S. & Burton, W. B. *Astrophys. J.* **226**, 790-816 (1978).
41. Nguyen-Q-Rieu *et al. Astr. Astrophys.* **46**, 413-428 (1976).
42. Guibert, J., Elitzur, M. & Nguyen-Q-Rieu *Astr. Astrophys.* **66**, 395-405 (1978).
43. Bujarrabal, V. *et al. Astr. Astrophys.* **81**, 1-7 (1980); **91**, 283-289 (1980).
44. van der Kruit, P. C. *Astr. Astrophys.* **15**, 110-122 (1971).
45. Hummel, E., van der Hulst, J. M. & Dickey, J. M. *Astr. Astrophys.* **134**, 207-221 (1984).
46. Weedman, D. W. *Astrophys. J.* **266**, 479-484 (1983).
47. Soifer, B. T. *et al. Astrophys. J. Lett.* **278**, L71 (1984).
48. Kennicutt, R. C. & Keel, W. *Astrophys. J.* **279**, L5-L9 (1984).
49. Lonsdale, C. J., Persson, S. E. & Matthews, K. *Astrophys. J.* (in the press).
50. Whiteoak, J. B. & Gardner, F. F. *Astrophys. Lett.* **15**, 211-215 (1973).
51. Baan, W. A. & Haschick, A. D. *Astr. J.* **88**, 1088-1093 (1983).
52. Moorwood, A. F. & Glass, I. S. *Astr. Astrophys.* **135**, 281-288 (1984).
53. IRAS bulletin 1 *Nature* **303**, 480 (1983); IRAS bulletin 15 *Nature* **310**, 548 (1984).
54. Rieke, G. H. & Lebovsky, M. J. *Astrophys. J. Lett.* **220**, L37-L41 (1978).
55. Whiteoak, J. B. & Gardner, F. F. *Astrophys. J. Lett.* **195**, L81-L84 (1975).
56. Condon, J. J., Condon, M. A., Gisler, G. & Puschell, J. J. *Astrophys. J.* **252**, 102-124 (1982).
57. Kronberg, P. P., Biermann, P. & Schwab, F. F. *Astrophys. J.* **246**, 751-760 (1981).
58. de Bruyn, A. G. *Astr. Astrophys.* **58**, 221-236 (1977).
59. Van Albada, G. D. thesis, Univ. Leiden (1978).
60. Wilson, A. S. & Ulvestad, J. S. *Astrophys. J.* **275**, 8-15 (1983).
61. Duric, N. *et al. Astrophys. J.* **273**, L11-L15 (1983).
62. Shostak, S. *Astr. Astrophys.* **68**, 321-341 (1978).
63. Gottesman, S. *et al. Astrophys. J.* **204**, 699-702 (1976).
64. Combes, F., Gottesman, S. & Weliachen, L. *Astr. Astrophys.* **59**, 181-187 (1977).

Sources of acidification in Central Europe estimated from elemental budgets in small basins

Tomas Paces*

Geological Survey of Prague, Malostranske nam. 19, 118 21 Praha 1, Czechoslovakia

Long-term increase in acidification of water and soils in the Elbe river basin is caused mainly by the deposition of dry SO₂ and the application of industrial fertilizers. The direct input of hydrogen ions by acid rain is less significant. The budgets of elements and protons indicate natural and anthropogenic processes that acidify and buffer our environment.

CENTRAL Europe is a region seriously damaged by acid industrial emissions, as demonstrated by increased rates of corrosion^{1,2}, deterioration in forest productivity^{3,4} and release of elements such as aluminium⁵ and heavy metals⁶ in water. A large-scale application of industrial fertilizers causes acidification of arable land and eutrophication of surface water^{7,8}. Central Europe receives one of the highest inputs of sulphur from the atmosphere⁹; the runoff of anions of strong acids in surface water has increased while the runoff of weak acids, represented by alkalinity of stream water, has decreased in the past century¹⁰. The data in Table 1 present convincing evidence, independent of uncertain historical measurements of pH, that acidification has occurred. It may be related to acid rain¹¹, deposition of dry gaseous SO₂ and NO_x from the atmosphere on plants and soils¹², intensive application of industrial fertilizers^{13,14} and depletion of organic matter from topsoil. These processes are of anthropogenic origin and affect the biogeochemical cycles and mass balances of naturally occurring chemical elements.

To collect quantitative data on the environmental changes and their rates, I have evaluated and compared mass budgets of H⁺, Na, K, L, Mg, Ca, Fe, Al, Si, N-NH₄⁺, N-NO₃⁻, C-HCO₃⁻, Cl, S and P in small hydrologically and geologically well-defined drainage basins in the watershed of the Elbe river (Fig. 1). Four representative basins considered here have similar bedrock, soil and climate conditions but differ in mean slopes and anthropogenic impacts. Comparison of their elemental budgets will help to elucidate the mechanisms that affect the fluxes of elements and cause acidification of soil and water. The data are

used to calculate weathering and erosion rates influenced by acid emissions and agricultural practice. The elemental budgets in basins affected by industry and agriculture serve as background data for a comparison with representative basins in other regions.

Regional characteristics

Representative basins shown in Fig. 1 are located on highlands composed of gneisses with interlayers of quartzites; the bedrock primarily contains oligoclase, quartz and biotite with <5% muscovite, orthoclase and sillimanite. Gneisses are covered by colluvium with well-developed soil profiles to a depth of 120 cm. The soils are typical brown-to-acid brown earths (Food and Agricultural Organization classification: Eutric Cambisol to Distric Cambisol gleyic). Elevation, mean slopes of terrain and major hydrological characteristics of the basins are summarized in Table 2. The mean annual temperature varies from 5 to 7 °C. Two of the basins, X-0 and X-8, are covered with coniferous forest consisting mainly of spruce. The coniferous part of the

Table 1 Long-term change in the runoff of anions from Central Europe, using the Elbe river at Lovosice, Czechoslovakia¹⁰

Year	1892	1976	Mean annual change in water concentration
	(mmol m ⁻² yr ⁻¹)		(μmol l ⁻¹ yr ⁻¹)
Cl ⁻	47	177	10
NO ₃ ⁻	5.6	40	2.7
SO ₄ ²⁻	36	170	11
HCO ₃ ⁻	360	252	-4.3

* Present address: Department of Land Improvement and Drainage, Royal Institute of Technology, S-100 44 Stockholm, Sweden.

forest in basin X-14 died several years before our observations started in 1975, but ~30% of the basin is still covered with deciduous trees and bushes, mainly oaks and beeches. Wheat, rye, barley, oats, maize, potatoes, clover, grass, flax and mustard seeds are rotated crops grown in basin X-7. These fields are intensively fertilized with industrial fertilizers and smaller amounts of organic matter, whereas the forested basins are not fertilized.

Basins X-0, X-7 and X-8 are located in a rural region of the Trnavka river basin (Pacov, Czechoslovakia) without local sources of industrial pollution. The mean concentration of SO_2 in air calculated from daily measurements in 1983 was $8.3 \mu\text{g m}^{-3}$, with a monthly variation coefficient of 37%. The concentration of NO_x was $11.2 \mu\text{g m}^{-3}$, with a variation coefficient of 31% (ref. 15). The forested basin X-14 is on the eastern slope of the Ertzgebirge Mountains facing the industrial region of the city of Most with open-pit coal mines and power plants in which local coal with a high sulphur content is burned. The 3-yr mean concentration of SO_2 in air was $112 \mu\text{g m}^{-3}$ with a variation of 52%, whereas the mean concentration of NO_x was $33 \mu\text{g m}^{-3}$ with a summer and winter half-year variation coefficient of 38% (ref. 15).

I have grouped the representative basins in three pairs to illustrate the influence of various factors on their elemental budgets, the first basin of each pair serving as a reference, the second showing the influence of a given factor. The effect of mean slope in forest is demonstrated by the pair X-0 and X-8; both basins are located in the rural region and have slopes of 3.8 and 13%, respectively. The effect of acid industrial emissions in forests is shown by the pair X-8 and X-14 and the effect of modern agriculture by the pair X-0 and X-7. The comparison is based on annual arithmetic means for observation period 1976–82 (7 hydrological years) in basins X-0, X-8 and X-7 and for observation period 1978–82 (5 hydrological years) in basin X-14.

Input and output mechanisms

Atmospheric precipitations were measured at 10 stations by international meteorological methods. Samples for chemical analysis were collected 1.3 m above ground, rainwater through glass funnels in polyethylene bottles and snow directly in polyethylene buckets. The diameters of the collectors were 25 and 30 cm. The containers were exposed only during individual precipitation episodes to determine the composition of wet precipitation. A similar set of collectors was exposed permanently for approximate estimation of total deposition. Samples were stored in a refrigerator immediately after each precipitation episode and analysed once a month. They were not filtered or acidified before chemical analysis^{16,17}.



Fig. 1 Location of representative basins in the watershed of the Elbe river. Numbers and symbols correspond to the system established by the Geological Survey of Prague.

Runoff water from the basins was measured continuously at gauge stations in stream channels. The subsurface discharge was found to be negligible by a detailed hydrological and geophysical investigation¹⁸. Samples for chemical analysis were collected once a month, but more frequently during extreme hydrological episodes such as spring thaw and summer rainstorms. This frequency of sampling was found to be satisfactory because the variation in electric conductivity of the set of collected samples was not statistically different from the variation in the conductivity measured in runoff every day. The samples were filtered through a $0.45\text{-}\mu\text{m}$ filter within 1 day of collection. In the field, water was filtered through a $0.1\text{-}\mu\text{m}$ filter for determination of dissolved Al. This method of sampling gives reproducible results because all sampled streams are very low in concentration of organic matter³⁰. Blank samples of deionized water were prepared and analysed by the same methods to check for possible contamination during sampling and analytical procedures. The pH of water was measured potentiometrically in the field; the flux of dissolved elements in water was calculated by multiplying the discharge of water between two sampling intervals and the arithmetic mean of the concentrations.

Biological output from the forested basins was determined from detailed records of biomass production in individual plots of the forests and chemical analyses of randomly selected trees. Mean annual production of biomass was multiplied by the mean concentration of elements in fresh wood to estimate the output of biologically fixed elements in basins X-0 and X-8. No data on biological fixation are available for basin X-14 because of

Table 2 Major characteristics of the representative basins in the watershed of the Elbe river

No. of basin	X-0	X-8	X-7	X-14
Locality	Hartvikov	Salacova Lhota	Vocadlo at Velky Jezov	Vysoka Pec at Most
Region	Ceskomoravska vysocina highland			Krusne hory (Ertzgebirge) Mountains
Type of region	Rural	Rural	Rural	Industrial
Type of basin	Forest	Forest	Field	Damaged forest
Drainage area (km^2)	0.98	1.68	0.59	2.7
Forested area (%)	100	100	1.3	30
Arable area (%)	0	0	98.7	0
Clearcut area (%)	0	0	0	70
Elevation above sea level				
Maximum (m)	724	744	635	924
Minimum (m)	672	557	570	335
Mean slope (%)	3.8	13.3	4.9	18.0
Annual precipitations (mm)	781	685	736	812
Annual discharge (mm)	108	128	171	431
pH of precipitations	4.27	4.27	4.27	4.19
pH of runoff	6.84	7.14	6.15	4.90

Table 3 Precipitation, P_i , and runoff, R_i , in representative basins

No. of basin	X-0		X-8		X-7		X-14	
Type of region	Rural		Rural		Rural		Industrial	
Type of basin	Forest		Forest		Field		Damaged forest	
Type of flux	P_i	R_i	P_i	R_i	P_i	R_i	P_i	R_i
H ⁺	0.42	0.000	0.37	0.000	0.40	0.001	0.46	0.054
Na ⁺	0.79	6.6	0.69	7.0	0.75	17	2.2	22.5
K ⁺	0.54	1.5	0.46	1.9	0.51	7.4	1.7	6.8
Mg ²⁺	0.38	2.6	0.33	3.8	0.36	16	1.1	26
Ca ²⁺	3.4	5.9	2.9	9.9	3.2	49	7.4	74
Fe ²⁺	0.16	0.18	0.14	0.19	0.15	0.33	0.15	0.20
Al	0.79*	0.22	0.68*	0.11	0.74*	0.06	1.0*	3.2
Si	0.7	14.4	0.7	15.2	0.7	15.0	0.7	33
Cl ⁻	2.4	3.2	2.0	4.3	2.2	43	5.1	16
N-NO ₃ ⁻	4.3	0.56	3.7	0.58	4.0	19	5.5	12
N-NH ₄ ⁺	5.6	0.00	4.9	0.00	5.3	0.00	7.5	0.00
S-SO ₄ ²⁻	12.2	8.0	10.5	9.0	12	28	19.6	96
P-P ₂ PO ₄ ⁻	0.09	0.025	0.075	0.024	0.08	0.023	0.08	0.042
C-HCO ₃ ⁻	0.0	2.9	0.0	6.3	0.0	5.7	0.0	0.0

The values are mean annual fluxes in kg ha⁻¹ yr⁻¹; means represent 7 hydrological years in basins X-0, X-8 and X-7 and 5 hydrological years in basin X-14.

* The high atmospheric input of aqueous Al probably results from the hydrolysis of dust in the atmosphere. This hydrolysis buffers the pH of rain before it falls in the basins. This is reflected by the lower values of the precipitation flux of hydrogen ions, P_{H^+} .

dieback of 70% of the tree population. I estimate the biological fixation in basin X-14 to be about one-third of the fixation in basin X-8. The data on biological fixation represent only the long-term output resulting from lumbering. I did not obtain reliable data on the net accumulation of biomass in the basins, therefore, I assume that the biomass remaining in the basins is recycled and that the net annual accumulation is approximately equal to the annual output by lumbering.

Anthropogenic input in field basin X-7 includes fertilization and sowing. The quantities of the applied fertilizers and seeds were recorded and samples collected. The applied quantity multiplied by mean concentrations of elements in fertilizers and seeds was summed over individual hydrological years to represent annual anthropogenic input. Biological output from basin X-7 was determined by recording the harvested amounts of crops and analysing samples from individual fields. Grains and straw, tubers and haulm were analysed separately. Water, rocks and soils were analysed in our laboratory by atomic absorption spectroscopy for Na, K, Mg, Ca and Fe and atomic absorption flameless spectroscopy for Al and photometry for Si. NH₄⁺ and Cl⁻ ions were determined by ion-sensitive electrodes and NO₃⁻ and SO₄²⁻ by standard methods of wet analysis¹⁹. Organic samples were analysed by a specialized agricultural laboratory (District Agricultural Testing and Control Station, Brno). Precision of the analyses of water was tested against international standards²⁰ with good results. The precision of the chemical analyses of organic samples and fertilizers was not reported. The accuracy of the analyses, however, is probably greater than my determination of the mass of organic matter entering and leaving the basins. The accuracy is 7% for precipitation and runoff water analyses and 30% for the anthropogenic input and biological output measurements.

Budgets of elements

The fluxes of elements resulting from wet precipitation and runoff from the basins are summarized in Table 3; anthropogenic input and biological output are given in Table 4.

Inputs, outputs and accumulation affected by acidification but which cannot be easily measured in field include: (1) input of elements by total (that is, chemical plus mechanical) weathering of bedrock; (2) output by chemical erosion; (3) output by mechanical erosion; (4) accumulation of sulphur in regolith; and (5) production and consumption of hydrogen ions (proton balance). These fluxes are estimated with a help of various mass-balance models.

Rate of total weathering of bedrock, rates of chemical and mechanical erosion and related inputs and outputs of elements

are calculated using a steady-state model of weathering^{10,21} (assuming that Na and Si in regolith are in a steady state). The annual accumulation or depletion of these elements in weathered zone is negligible compared with their annual inputs and outputs so that the amounts in regolith of the basins are approximately constant with time. In such a case, the input of an element i resulting from total weathering of bedrock, W_i (kg per hectare (ha) per yr), can be calculated from a steady-state mass-balance equation

$$W_i = -(P_i - R_i + A_i - B_i - M_i) \quad (1)$$

where P_i is input of chemical element i by atmospheric precipitation, R_i is output of element i by runoff from drainage basin, A_i is anthropogenic input of element i , B_i is output of biologically fixed element i and M_i is output of element i by mechanical erosion of weathered material.

Two equations (1) written for $i = \text{Na}$ and Si contain four unknown values: W_{Na} , W_{Si} , M_{Na} and M_{Si} . These values are related to a rate of total weathering of bedrock, W_{rock} , and to mechanical erosion of regolith, M_{rghth} , by the following equations

$$W_i = W_{\text{rock}} x_{i,\text{rock}} \quad (2)$$

$$M_i = M_{\text{rghth}} x_{i,\text{rghth}} \quad (3)$$

Table 4 Biological output, B_i , and anthropogenic input, A_i , in representative basins

No. of basin	X-0	X-8	X-7	
Type of region	Rural	Rural	Rural	
Type of basin	Forest	Forest	Field	
Type of flux	B_i	B_i	A_i	B_i
Na ⁺	0.24	0.28	2.3	0.5
K ⁺	3.4	3.9	84.9	110
Mg ²⁺	0.95	1.1	3.5	7.4
Ca ²⁺	7.6	9.4	64.4	23.5
Fe ²⁺	0.26	0.30		
Cl ⁻	0.6	0.6	45.9	2
S-SO ₄ ²⁻	0.56	0.61	18.3	2
N-NH ₄ ⁺			57.6	
N-NO ₃ ⁻			51.0	
N	7.6	8.7		106
P-H ₂ PO ₄ ⁻	0.84	0.96	36.2	15

Values are mean annual fluxes, kg ha⁻¹ yr⁻¹.

where $x_{i,rock}$ and $x_{i,rglth}$ are dimensionless concentrations ($g\ g^{-1}$) of element i in bedrock and its regolith respectively. Including W_{rock} and M_{rglth} , there are six unknown and six independent equations (1), (2) and (3) written for $i = Na$ and Si . This system of equations is solved to yield the rate of mechanical erosion of regolith

$$M_{rglth} = \{[(P_{Na} - R_{Na} + A_{Na} - B_{Na})/x_{Na,rock}] - [(P_{Si} - R_{Si} + A_{Si} - B_{Si})/x_{Si,rock}]\} / [(x_{Na,rglth}/x_{Na,rock}) - (x_{Si,rglth}/x_{Si,rock})] \quad (4)$$

and mechanical erosion of any element, M_i , according to equation (3). Substituting the calculated values of mechanical erosion of Na and Si , M_{Na} and M_{Si} in equation (1), we obtain the rate of the input of the elements resulting from total weathering of bedrock, W_{Na} and W_{Si} . Finally, either of the last-mentioned values substituted in equation (2) yields the rate of total weathering of bedrock, W_{rock} (Table 5).

The rate of chemical erosion of weathered bedrock, C_{rock} , is calculated from equation (5)

$$C_{rock} = -\sum_i (P_i - R_i + A_i - B_i) m_{oxide}/n.m_i \quad (5)$$

where $\sum$ includes all elements composing bedrock except O , m_{oxide} is the relative molecular mass of the respective oxide, n is the stoichiometric coefficient of the cation in the oxide (for example, $n=2$ for Na_2O) and m_i is atomic mass of the cation i . The flux resulting from chemical erosion of weathered bedrock includes direct dissolution of bedrock (that is, chemical weathering) and desorption of exchangeable ions from soil. Mechanical erosion of bedrock is defined by equation

$$M_{rock} = W_{rock} - C_{rock} \quad (6)$$

Note that the rate of mechanical erosion of bedrock, M_{rock} , is different from the mechanical erosion of regolith, M_{rglth} , calculated from equation (4), because weathered material contains components not only derived from bedrock but also others, such as water molecules and organic matter. The calculated rates of chemical and mechanical erosion of gneisses in the basins examined here are summarized in Table 5. The effect of various factors on weathering and erosion are expressed by ratios of the characteristic rates for the individual basins.

Accumulation of S in regolith of the basins, which is an important process of environmental acidification, is expressed by a mass-balance equation

$$P_S - R_S + A_S - B_S + W_S + D_S + G_S = \sum_S \quad (7)$$

where $\sum_S$ is accumulation of S in $kg\ ha^{-1}\ yr^{-1}$ (this value contains all the uncertainties and errors inherent in the values of fluxes in equation (7)); fluxes P_S , R_S , A_S , B_S were measured in the field, W_S was calculated from the weathering rate of bedrock, W_{rock} , in Table 5 and mean concentration of S in bedrock from equation (2); D_S is the atmospheric input of S during dry periods calculated as a difference between the values measured by permanently exposed collectors and by collectors exposed only to wet precipitations; G_S is deposition of dry $S-SO_2$ from the atmosphere. A small portion of D_S may result from the deposition of SO_2 in the permanently exposed collectors, but this is negligible compared with the deposition of SO_2 on vegetation and soil (G_S).

The deposition of dry SO_2 was estimated from the mean deposition velocity and the mean concentration of SO_2 in the local atmosphere. Deposition velocity was calculated from published estimates for snow cover and for dry and wet periods during which stomata of trees were opened and closed²². The deposition velocity in forested basins X-0 and X-8 of the rural region is $5.8\ mm\ s^{-1}$. The mean concentration of SO_2 is $8.3\ \mu g\ m^{-3}$ and the corresponding deposition rate of dry SO_2 is

Table 5 Rates of total weathering, chemical erosion and mechanical erosion of gneisses

No. of basin Type of region Type of basin	X-0 Rural Forest	X-8 Rural Forest	X-7 Rural Field	X-14 Industrial Damaged forest
Total weathering, W_{rock}^*	420	732	1,700	1,510
Chemical erosion, $C_{rock}^\dagger$	64	109	160	318
Mechanical erosion, $M_{rock}^\dagger$	356	623	1,540	1,192
Determining factors				
Ratio	X-8/X-0	X-7/X-0	X-14/X-8	
Wet precipitations	0.88	0.94	1.2	
Slope of terrain	3.4	1.3	1.4	
Concentration of SO_2 in air	1	1	13	
Fertilization	no/no	yes/no	no/no	
Response factors				
Effect of	Slope	Agriculture	Emissions	
Ratio	X-8/X-0	X-7/X-0	X-14/X-8	
W_{rock}	1.74	4.0	2.1	
C_{rock}	1.71	2.5	2.9	
M_{rock}	1.88	4.3	1.9	

* Calculated according to equations (1), (2), (3) and (4) using fluxes of Na and Si in Tables 3 and 4 and Na and Si mean concentrations ($g\ g^{-1}$) in bedrock and regolith in the individual basins ($x_{Na,rock}$; $x_{Na,rglth}$; $x_{Si,rock}$; $x_{Si,rglth}$): X-0 (0.022; 0.0085; 0.30; 0.30), X-8 (0.016; 0.0070; 0.30; 0.28), X-7 (0.016; 0.0067; 0.31; 0.27), X-14 (0.02; 0.007; 0.3; 0.3).

† Calculated according to equations (5) and (6). Units of rates are $kg\ ha^{-1}\ yr^{-1}$ of bedrock. Ratios are dimensionless.

Table 6 Budget of S in representative basins

No. of basin Type of region Type of basin	X-0 Rural Forest	X-8 Rural Forest	X-7 Rural Field	X-14 Industrial Damaged forest
Wet precipitations, P_S	12.0	10.5	12.0	19.6
Runoff, $-R_S$	-8.0	-9.0	-28.0	-96.0
Anthropogenic input, A_S	0	0	13.3	0
Biological output, $-B_S$	-0.56	-0.61	-2.0	-0.3
Weathering of bedrock, W_S^*	0.70	1.2	2.9	3.8
Deposition during dry periods, D_S	1.0	3.3	3.0	0.2
Deposition of gaseous SO_2 , G_S	7.6	7.6	7.6	88.6
Accumulation in basins, $\sum(S)$	12.7	13.0	13.8	15.9

* Calculated from the rate of weathering of bedrock (W_{rock}) and mean concentrations of S in gneiss, $x_{S,rock}$ (0.0017 in X-0, X-8, X-7 and 0.0025 in X-14) according to equation (2). It is assumed that all S from weathered bedrock dissolves and none is preserved as residual sulphides. Units, $kg\ ha^{-1}\ yr^{-1}$ sulphur.

$7.6\ kg\ ha^{-1}\ yr^{-1}$ of S . I do not have adequate data on the deposition velocity on fields in basin X-7, but assume that it is the same as the forest basins of the rural region. The deposition velocity may be slightly lower in basin X-14 because of the longer duration of snow cover at the higher altitude and the thin biological cover. The velocity is estimated to be $5\ mm\ s^{-1}$ and the mean concentration of SO_2 $112\ \mu g\ m^{-3}$, with the corresponding flux $88.6\ kg\ ha^{-1}\ yr^{-1}$ of S (only one digit of the value is significant). The fluxes of S and calculated accumulation according to equation (7) are given in Table 6. I find that the accumulation is approximately the same in all the basins despite the large difference in the concentration of SO_2 in the air of the rural and industrial regions.

Table 7 Hydrogen-ion production and consumption by known geochemical, biological and anthropogenic processes

No. of basin Type of region Type of basin	X-0 Rural Forest	X-8 Rural Forest	X-7 Rural Field	X-14 Industrial Damaged forest
Input of H_3O^+ by wet precipitation	42.2	37.3	39.3	45.7
Output of H_3O^+ by runoff	-0.02	-0.01	-0.14	-5.4
Hydrolysis reactions involving alumina; flux calculated from P_{Al} , R_{Al} , pH of precipitation, pH of runoff and complexing constants for Al-hydroxy-complexes ²⁹	8.8	7.3	7.5	-13.0
Oxidation of pyrite calculated from the rate of release of S from bedrock, W_{S} (Table 6), according to stoichiometry $\text{FeS}_2 + 3.5 \text{O}_2 + \text{H}_2\text{O} = \text{Fe}^{2+} + 2 \text{SO}_4^{2-} + 2\text{H}^+$	2.2	3.9	9.0	11.8
Precipitation of ferric hydroxide calculated from inputs and outputs of Fe^{2+} and the amount of Fe^{2+} released from oxidation of pyrite $\text{Fe}^{2+} + 0.25 \text{O}_2 + 2.5 \text{H}_2\text{O} = \text{Fe}(\text{OH})_3 + \text{SH}^+$	2.1	3.7	8.7	11.6
Dissolution and dissociation of CO_2 according to stoichiometry $\text{CO}_2 + \text{H}_2\text{O} = \text{HCO}_3^- + \text{H}^+$	23.9	52.5	47.5	0.0
Biological fixation of cations and anions (except nitrogen species) in biomass removed by harvest and lumbering, calculated from data in Table 4	51.4	59.6	232.3	(20)*
Transformation of nitrogen species according to stoichiometry $\text{NH}_4^+ + \text{ROH} = \text{RNH}_2 + \text{H}_2\text{O} + \text{H}^+$; $\text{NO}_3^- + \text{ROH} + \text{H}^+ = \text{RNH}_2 + 2\text{O}_2$; $\text{NH}_4^+ + 2\text{O}_2 = \text{NO}_3^- + \text{H}_2\text{O} + 2\text{H}^+$ where R is organic matter	13.3	12.7	192.1	100
Atmospheric deposition of dry SO_2 and oxidation according to stoichiometry $\text{SO}_2 + \text{H}_2\text{O} + 0.5 \text{O}_2 = \text{SO}_4^{2-} + 2\text{H}^+$	47.5	47.5	(50)*	554
Release of cations and anions by hydrolysis of bedrock and depletion of exchangeable cations in soil calculated from budgets of Na, K, Mg, Ca, Cl and P in Tables 3, 4	-111	-151	-411	-609
Σ ('excess' production of protons)	80	74	175	116
Deviation from electrical neutrality of aqueous inputs and outputs: $\Sigma(\text{cations}) - \Sigma(\text{anions})$ in precipitations minus $\Sigma(\text{cations}) - \Sigma(\text{anions})$ in runoff calculated from $\text{meq m}^{-2} \text{yr}^{-1}$	6.0	12	7.9	4.3

Units, $\text{mmol m}^{-2} \text{yr}^{-1}$ of H^+ ; * uncertain values (see text).

The process of acidification is best represented by the proton balance of all fluxes and reactions which involve an exchange of H^+ such as acid rain, ion exchange, hydrolysis of minerals, biological fixation of ions, oxidation of pyrite and SO_2 in regolith and nitrification²³⁻²⁵. The inputs, outputs, production and consumption of H^+ in $\text{mmol m}^{-2} \text{yr}^{-1}$ calculated from the measured and estimated fluxes of elements in the representative basins are shown in Table 7. The most important source of H^+ is not acid rain but the deposition and subsequent oxidation of dry SO_2 , transformation of nitrogen species and biological fixation of cations. The most important sink of H^+ is hydrolysis of rock-forming minerals and ion exchange in soil. Taking into account all the known processes that involve the exchange of protons, there is an 'excess' production of H^+ in all the representative basins (Table 7). This 'excess' production is not an artefact of electrical imbalance of chemical analyses (Table 7); it must be consumed in the soil by processes that are not apparent in the considered fluxes, probably the protonation of organic active groups in soils, specific adsorption of H^+ on negative surfaces of clay minerals and sesquioxides and hydrolysis of dead organic matter. These processes, together with the depletion of exchangeable cations, gradually exhaust the buffering capacity of local soils.

Discussion

Acidification of soil and water is a dynamic process which depends on fluxes of acidifying chemicals and on geochemical and biochemical reactions involving an exchange of protons in the whole ecosystem. Harmful effects such as the die-back of trees occur when the excess production of protons lasts long enough to exhaust the buffering capacity of soils. The causes of acidification in Central Europe are natural and anthropogenic, the major natural process being the biological fixation of cations and the major anthropogenic processes the emissions of SO_2 and the application of industrial fertilizers. The direct input of H^+ by acid rain is of importance in the forests of rural region while in the field basin it is secondary to biological acidification and in the industrial region it is less important than the input

of H^+ resulting from deposition and oxidation of dry SO_2 . Geochemical processes such as oxidation of sulphide minerals and precipitation of ferric hydroxide are not important.

The rural region with a relatively low concentration of SO_2 in air and the industrial region with a very high concentration of SO_2 are 160 km apart. The average concentration gradient of SO_2 in air is very steep, $0.65 \mu\text{g m}^{-3} \text{km}^{-1}$. It corresponds to the increase in H^+ production by the oxidation of adsorbed SO_2 , equal to $3.2 \text{mmol m}^{-2} \text{yr}^{-1} \text{km}^{-1}$ from the point-source of emissions. This is in contrast to a very small increase in the input of H^+ resulting from acid rain which is only $\sim 0.06 \text{mmol m}^{-2} \text{yr}^{-1} \text{km}^{-1}$. This gradient in the input of SO_2 correlates with the observed die-back of coniferous trees and the acceleration of weathering and erosion accompanied by the increased output of nutrients. The total weathering rate of gneisses increases by a factor of 2.1, chemical erosion by 2.9 and mechanical erosion by 1.9 in forest in the industrial region. Part of the increase is caused by the steeper slope of basin X-14 and by higher precipitation and discharge rates. The runoff of all the elements (except Fe) increases (Table 3). The decrease in pH of runoff from basin X-14 causes the alkalinity of water to fall below the detection limit so that water in the basin loses its carbonate buffering capacity. At the same time, alumina starts to behave as a buffer and consumes protons at the rate of $13 \text{mmol m}^{-2} \text{yr}^{-1}$ (Table 7). This leads to a large increase in the runoff of dissolved Al from $0.11 \text{kg ha}^{-1} \text{yr}^{-1}$ in basin X-8 to $3.2 \text{kg ha}^{-1} \text{yr}^{-1}$ in basin X-14 (Table 3). Acidification in the forest of the industrial region is accompanied by a large increase in runoff of nitrates from $0.58 \text{kg ha}^{-1} \text{yr}^{-1}$ of N in basin X-8 to $12 \text{kg ha}^{-1} \text{yr}^{-1}$ in basin X-14. The increased flux of nitrates does not originate from polluted air because the mean annual concentration of NO_x in air increases three times (from 11 to $33 \mu\text{g m}^{-3}$) while the output of nitrates increases >20 times. Therefore, the increased output reflects the decreasing denitrification capacity of damaged forest in basin X-14. This agrees well with the increased runoff of nitrates from a similar forest basin in New Hampshire which was clear cut experimentally²⁶.

The consumption of H^+ by ion exchange in soil and hydrolysis of bedrock is not fast enough to neutralize the acidifying inputs

in any of the forest basins described here. The excess production of 80 and 74 mmol m⁻² yr⁻¹ in the forests of basins X-0 and X-8 does not seem to have a damaging effect on trees; it does not lower runoff pH and does not increase the output of dissolved Al and nitrates. But the buffering capacity of local soils may soon be exhausted. It is possible that the excess hydrogen ions protonate organic active groups and that they are specifically adsorbed on clay minerals and sesquioxides, causing a reversal in surface charge of some soil particles²⁷. If the specific adsorption of excess H⁺ is irreversible, then the exhaustion of the buffering capacity of soils and the changes in the surface charge of soil particles may have a sudden negative effect on the productivity of the forest.

The studies reported here support the suggestion²⁸ that modern agriculture contributes to acidification. The major acidifying mechanism is the application of fertilizers containing the weak base NH₃ and strong acids H₂SO₄ and HNO₃. Agricultural practices in basin X-7 speed up the total weathering of gneiss by factor of four, but most of the weathered bedrock is removed by mechanical rather than chemical erosion. The increased runoff of dissolved substances is caused by fertilization; ~89% Cl, 83% S and 17% N introduced in fertilizers are leached out. On the other hand, cations are retained by plants contributing to acidification of soil and water. The pH of runoff from field basin X-7 decreases by ~1 pH unit to pH 6.15, compared with the runoff from its reference basin X-0. The decrease in pH is not enough to mobilize Al, whose flux remains low. The excess production of protons in basin X-7 is 175 mmol m⁻² yr⁻¹, 95 mmol m⁻² yr⁻¹ more than in the reference basin X-0. Although the uncertainty in this value is large because of the less precise values of biological fixation, the general acidifying trend in the field basin is clear.

The observed differences in the fluxes between the basins representing the agricultural and industrial influence on cycling of elements are not affected greatly by different slopes of terrain. Although the ratio of slopes X-8/X-0 is 3.4, the ratio of chemical erosion is only 1.7. The ratio of slopes X-14/X-7 is 3.7 and the ratio of chemical erosion 2.0, but basin X-14 receives more atmospheric precipitations and the runoff is faster. The hydrological factor increases input and output rates of elements including H⁺ and is therefore an important factor of acidification. The fast rate of flushing through basin X-14 may be the cause of the relatively small accumulation of S in spite of an extremely high input of dry SO₂. The similar accumulation rate of S in the same type of soils in the rural and industrial regions indicates that the properties of soils may be important in S accumulation.

The measured and calculated fluxes of chemical elements in the representative basins indicate that observed long-term trends in the flux of dissolved elements from the Elbe river in central Europe are caused mainly by industrial emissions and intensive application of inorganic fertilizers. The runoff from central Europe is still effectively buffered because the mean pH of the Elbe river is 7.1. Locally, however, the pH of runoff drops below the capacity of CO₂/HCO₃ buffer; here, Al is mobilized and its runoff increased. In the areas with strong input of dry SO₂ and intensive application of inorganic fertilizers, the soils are acidified even if the acidification of runoff is small. Acidification can gradually exhaust the buffering capacity of soils and change surface charge on soil particles. At an 'inflection' point, by analogy with acid-base titration, it may lead to a sudden drop in biological productivity and die-back of coniferous trees. This inflection point is not yet defined; its existence is suggested by the fact that the acidification of soils in Central Europe has proceeded for decades whereas the die-back of forests is a relatively fast phenomenon which takes only a few years.

The data described here are too preliminary to separate quantitatively the distribution of H⁺ ions that are generated in excess by natural and anthropogenic acidification. Some of the ions are consumed by ion exchange causing a depletion of the exchangeable cations in soil and loss of nutrients such as Ca²⁺ and Mg²⁺. Some H⁺ ions are used up by hydrolysis of rock-forming minerals; this process increases the rate of chemical weathering and may be ecologically beneficial because it releases nutrient cations from bedrock. Some of the produced H⁺ ions are consumed by organic acid groups and may be adsorbed specifically on negative surfaces of clay minerals and sesquioxides. This may reverse the charge of some active surface sites and negatively affect properties of soils. Further research is needed to clarify the distribution of excess H⁺ and the mechanisms of their consumption within water-soil-rock systems.

This paper was written at the Geological Institute of the Louis Pasteur University, Strasbourg. I thank Professor Y. Tardy for his support of this research. The data on precipitations were obtained by the research team of Dr B. Moldan and data on runoff, anthropogenic inputs and biological outputs were collected by Mrs E. Pacesova, all of the Geological Survey of Prague. I also thank Dr C. A. J. Appelo (Institute of Earth Sciences, Free University, Amsterdam), Dr N. van Breemen (Department of Soil Science & Geology, Agricultural University, Wageningen) and Dr H. M. Seip (Central Institute for Industrial Research, Oslo) for constructive criticism and valuable advice.

Received 23 October 1984; accepted 15 February 1985.

- Kucera, V. *Ambio* **5**, 243-248 (1976).
- Kulis, M. *Werkstoffe Korros.* **28**, 89-97 (1977).
- Ulrich, B. *Allg. Forst-Zeitschrift* **35**, 1198-1202 (1980).
- Ulrich, B. in *Ecological Effects of Acid Deposition*, 221-231 (Natl. Swedish environ. Protect. Bd Rep. PM 1636, 1983).
- Paces, T. in *Ecological Effects of Acid Deposition*, 161-173 (Natl. Swedish environ. Protect. Bd Rep. PM 1636, 1983).
- Kreutzer, K. in *Gewässerversauerung in der Bundesrepublik Deutschland* Materialien 1/84, 240-252 (Schmitt, Berlin, 1984).
- Prochazkova, L. in *Coupling of Land and Water Systems* (ed. Hasler, A. D.) 65-73 (Springer, Berlin, 1975).
- Prochazkova, L. *Vod. Hospod.* **B33**, 71-76 (1983).
- Dempier, W. *Ambio* **11**, 215-228 (1982).
- Paces, T. *Ambio* **11**, 206-208 (1982).
- Overrein, L. N., Seip, H. M. & Tollan, A. eds *SNSF Proj. Res. Rep. FR 19/80* (Norwegian Inst. Wat. Res., Oslo-Ås, 1981).
- Fowler, D. *Proc. int. Conf. Ecol. Impact of Acid Precip.* Sandefjord (eds Drablos, D. & Tollan, A.) 22-32 (SNSF Project-NLH, Ås, Norway, 1980).
- Skorepa, J., Vcislova, B. & Vrba, J. *Int. Symp. Impact. agric. Act. Ground Water*, Part 1, 313-322 (Int. Ass. Hydrogeol., Prague, 1982).
- Bertilsson, G. in *Acidification Today and Tomorrow*, 85 (Min. Agric., Stockholm, 1982).
- Czech Hydrometeorological Institute, Prague and Usti nad Labem (unpublished report, 1984).
- Moldan, B., Fottova, D. & Vesely, M. *Evaluation of Chemical Composition of Wet Precipitations in CSSR* (Arch. Ustredni Ustav Geologicky, Prague, 1982).
- Moldan, B. *Transport of Matter Through Atmosphere* (Archive Ustredni Ustav Geologicky, Prague, 1977).
- Paces, T. & Moldan, B. *Sh. geol. ved. Rada HIG* **15**, 157-195 (1981).
- Kobrova, M. et al. *Manual of Chemical Analyses of Natural Waters* (Ustredni ustav geologicky, Prague, 1983).
- Sixta, V. & Paces, T. *Vest. ustr. Úst. geol.* **53**, 225-231 (1978).
- Paces, T. in *Chemical Weathering of Rocks and Minerals* (eds Colman, S. M. & Dethier, D. P.) (Academic, New York, in the press).
- Hultberg, H., Grennfelt, P. & Olsson, B. *Wat. Sci. Tech.* **15**, 81-103 (1983).
- Driscoll, C. T. & Likens, G. E. *Tellus* **34**, 283-292 (1982).
- van Breemen, N., Mulder, J. & Driscoll, C. T. *Pl. Soil* **75**, 283-308 (1983).
- van Breemen, N., Driscoll, C. T. & Mulder, J. *Nature* **307**, 599-604 (1984).
- Borman, F. H. & Likens, G. E. *Scient. Am.* **223**, 92-101 (1970).
- Parks, G. A. *Adv. Chem. Ser.* **67**, 121-160 (1967).
- Acidification Today and Tomorrow*, 81-87 (Swedish Min. Agric. Envir. Committee, 1982).
- Hem, J. D., Roberson, C. E., Lind, C. J. & Polzer, W. L. *U.S. Geol. Surv. Wat. Supply Pap.* 1827-E (Washington, DC, 1973).
- Skorepa, J. *Int. Symp. Impact. agric. Act. Ground Water*, Part 1, 323-334 (Int. Ass. Hydrogeol., Prague, 1982).

Three-dimensional structure of calmodulin

Y. Sudhakar Babu^{*,†}, John S. Sack^{*,‡}, Trevor J. Greenhough^{*,†}, Charles E. Bugg^{*,†,||}, Anthony R. Means[¶] & William J. Cook^{*,†,§,¶}

Departments of Pathology[§] and Biochemistry^{||}, Comprehensive Cancer Center^{*} and Institute of Dental Research[†], University of Alabama at Birmingham, University Station, Birmingham, Alabama 35294, USA

[¶]Department of Cell Biology, Baylor College of Medicine, Houston, Texas 77030, USA

The three-dimensional structure of calmodulin has been determined crystallographically at 3.0 Å resolution. The molecule consists of two globular lobes connected by a long exposed α -helix. Each lobe binds two calcium ions through helix-loop-helix domains similar to those of other calcium-binding proteins. The long helix between the lobes may be involved in interactions of calmodulin with drugs and various proteins.

CALMODULIN is one member of a family of low relative molecular mass acidic calcium-binding proteins¹. Many of these proteins, such as parvalbumin, troponin-C, intestinal calcium-binding protein and S-100, are found only in vertebrates and, even in these animals, are restricted to a small number of tissues or cells. In addition, the function, when known, is specific and limited to one metabolic process or pathway. Calmodulin is the unique member of this protein class, as it is present in all eukaryotic cells and serves many functions^{2,3}. This Ca receptor is a single 148-amino-acid polypeptide chain containing four Ca-binding sites and is highly conserved throughout evolution³⁻⁶.

Although calmodulin exhibits significant α -helicity even in the absence of Ca, binding of this ion results in an increase in the α -helical content⁷⁻¹⁰. It is this Ca-induced conformational change that allows calmodulin to interact with enzymes, peptides and pharmacological agents such as the phenothiazines^{2,3}. A major question is how the protein can interact specifically with so many different molecules. It is possible that the regulation of different enzymes by calmodulin is partially a function of the number of Ca ions bound. However, controversy exists regarding the order in which the four sites are filled and the discrete structural changes that accompany Ca binding.

Here we describe the structure of calmodulin at 3.0 Å resolution (Fig. 1). Calmodulin represents the fourth member of the Ca-binding protein family to be characterized crystallographically. (The other three include parvalbumin¹¹, intestinal Ca-binding protein¹² and troponin-C^{13,14}.) However, it is the first ubiquitous intracellular Ca receptor to be subjected to such scrutiny and represents the first Ca-binding protein where the gene^{6,15} and protein structure can be compared directly.

Structure determination

Calmodulin from rat testes was isolated and purified as described previously⁷. The crystallization technique was the same as described by Cook and Sack¹⁶; all crystals were obtained from solutions of 2-methyl-2,4-pentanediol at pH 5.6. The crystals belong to triclinic space group P1 with unit cell dimensions $a = 29.84$ Å, $b = 53.72$ Å, $c = 24.94$ Å, $\alpha = 93.3^\circ$, $\beta = 97.45^\circ$ and $\gamma = 88.77^\circ$; there is one molecule per unit cell.

Two heavy-atom derivatives, lead acetate and potassium platinum tetrachloride, were used in the solution of the structure by the multiple isomorphous replacement method. They were obtained by soaking crystals in 75% (v/v) 2-methyl-2,4-pentanediol that contained 0.05 M cacodylate (pH 5.6) and the heavy-atom reagent. Many attempts to replace the Ca ions by lanthanide ions were unsuccessful.

Data for native and derivative crystals were collected with a Picker FACS-1 diffractometer at room temperature using an omega step-scan and nickel-filtered CuK α -radiation. Friedel mates were measured during data collection at negative 2θ

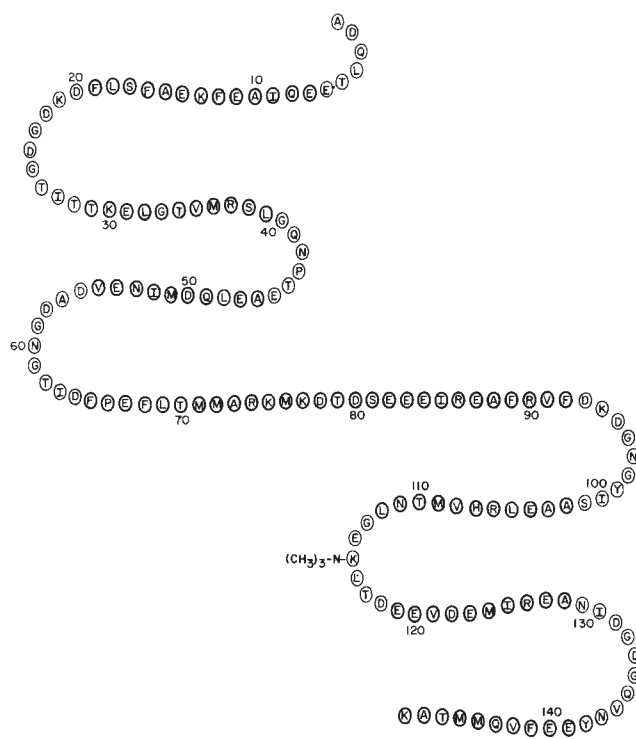


Fig. 1 Amino-acid sequence of calmodulin. The sequence is arranged to show schematically the topology of the molecule. Bold circles represent residues in α -helices.

values in groups of 10 reflections, alternating with their counterparts. Standard reflections were measured periodically and were used to correct for decomposition. Absorption corrections were applied according to the method of North *et al.*¹⁷.

The heavy-atom positions were determined by a combination of three-dimensional isomorphous difference Patterson and difference Fourier maps. Phases for the difference Fourier maps were calculated by the method of Blow and Crick¹⁸. Cross difference Fourier maps were used to bring the Pb and Pt coordinates to a common origin and, in conjunction with the use of anomalous data, to establish the correct enantiomeric arrangement of heavy atoms. The heavy-atom parameters were refined by the origin-removed Patterson method of Terwilliger and Eisenberg¹⁹. This procedure is based on the Patterson-function correlation method of Rossmann²⁰, except that the origins of the Patterson function are now removed from this correlation and centric and acentric reflections are treated separately. The final model for the Pb derivative consists of seven sites; two Pb ions substitute for Ca. The Pt derivative contains three sites. Various statistics for the heavy-atom derivatives are summarized in Table 1. The protein phases and the figures-of-merit were calculated by the method of Blow and Crick¹⁸.

[†] Present address: Department of Biochemistry, Rice University, Houston, Texas 77251, USA.

[¶] To whom correspondence should be addressed.

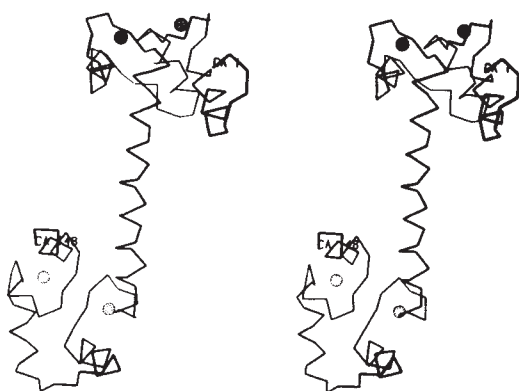


Fig. 2 Stereo drawing of the α -C backbone of calmodulin. The four Ca ions are represented by dotted circles.

Contributions from the anomalous differences were included in the phase calculations. The overall figure-of-merit is 0.71 for the complete data set to 3 Å resolution.

An electron density map at 3 Å resolution was computed from the 'best phases' with figure-of-merit weighting factors¹⁸. This map clearly showed the molecular boundary and most of the secondary structural features, but the polypeptide chain could not be followed completely without ambiguities. Sequence alignment was further complicated by the high degree of internal sequence homology. A modified electron density map was obtained by using a tentative chain tracing to assign a molecular boundary, flattening the solvent regions outside the molecular boundary and removing negative electron density values in the molecular boundary. This modified map was back-transformed to produce calculated phases, which were then combined with the multiple isomorphous replacement phases using the method of Bricogne²¹. A map calculated with the resultant phases resolved most of the ambiguities in chain tracing, primarily resulting from better definition of side-chain density; α -C atom positions for most of the residues were assigned from this map. A partial model was traced without residues 1–7, 75–80 and 129–148. These regions either had weak density or contained breaks in the density. The partial model was built and regularized using the computer graphics package FRODO²². Phases were calculated from this model and combined with the original multiple isomorphous replacement phases using the method of Bricogne²¹. The resultant map showed the protein density more clearly, allowing the model to be built completely. The four Ca positions were confirmed from a native anomalous difference Fourier map, calculated using the combined phases.

Structure description

A tracing of the α -C backbone of calmodulin is shown in Fig. 2. The molecule looks somewhat like a dumbbell with the two



Fig. 3 Stereo drawing of the two pairs of Ca-binding domains. The molecule has been split between residues 79 and 80 and the two halves shown in approximately the same orientation.

lobes connected by an eight-turn α -helix and no contacts between the two lobes. The length of the molecule is ~ 65 Å, with each lobe having approximate dimensions of $25 \times 20 \times 20$ Å. The two Ca-binding domains in each lobe are related to each other by an approximate 2-fold axis. The two individual lobes can be brought into approximate register by a rotation of $\sim 90^\circ$ coupled with a translation. The two halves of the molecule are remarkably similar; the corresponding α -carbon positions can be superimposed with an average difference of ~ 1 Å (Fig. 3). The molecule contains seven helices, I–VII, consisting of residues 7–19, 29–39, 46–55, 65–92, 102–112, 119–128 and 138–148. The helix content is 63%, which is about the same as observed in troponin-C^{13,14}. Calcium-binding domains 1 and 4 have a typical helix-loop-helix conformation and are similar to the two Ca-binding domains in parvalbumin¹¹ and to the III–IV domain in intestinal Ca-binding protein¹². Domains 2 and 3 also display helix-loop-helix conformations, but are different from that found in parvalbumin and intestinal Ca-binding protein in that these two domains share helix IV. This long helix provides both the carboxy-terminal helix for domain 2 and the amino-terminal helix for domain 3. Helix IV is two to three times longer than any of the other helices in the molecule. Because most of helix IV is exposed to solvent and forms only a few contacts with the rest of the molecule, it is possible that this part of the molecule retains some helical structure in the absence of Ca.

The Ca-binding segments consist of residues 20–31, 56–67, 93–104 and 129–140. These four segments are similar to the two that are described in parvalbumin¹¹ and to the calcium-binding region III–IV in intestinal Ca-binding protein¹². All of these contain 12 residues; the first 9 form the loop and the last 3 the beginning of the second helix in the helix-loop-helix domain. In all of these Ca-binding domains, the Ca ion is displaced from the centre of the loop towards the amino-terminal helix of the helix-loop-helix domain. The second helix of each domain orients the last residues of the Ca-binding segment so that they can participate in Ca coordination.

The molecule is stabilized by multiple interactions between the helices. In addition to these interactions, there is hydrogen bonding between adjacent Ca-binding loops in each half of the molecule. Residues 25–29 run antiparallel to residues 61–65 and these two stretches are joined by H bonds between Gly 25 and Asp 64 and between Ile 27 and Thr 62. In the second half of the molecule, the stretches 99–101 and 135–137 are also antiparallel and are joined by H bonds between Ile 100 and Val 136. Similar arrangements of the calcium-binding loops are found in parvalbumin and intestinal Ca-binding protein. Ca-binding domains in calmodulin are joined by linker segments composed of six residues; this compares with the linker segments of 8 residues in parvalbumin¹¹ and 10 in intestinal Ca-binding protein¹².

In each half of the molecule the two Ca ions are ~ 11.3 Å

Table 1 Statistics for the heavy-atom derivatives

	Pb(C ₂ H ₃ O ₂) ₂	K ₂ PtCl ₄
Resolution (Å)	3.0	3.2
Merging <i>R</i> (%)	4.7	2.5
Heavy-atom concentration (mM)	0.7	0.2
Fractional change (%)	30	12
No. of heavy-atom sites	7	3
<i>F_H</i> (calc)/ <i>E</i>	2.7	1.5

$$\text{Merging } R = \frac{\sum |F(h)| - |F(\bar{h})|}{\sum |F(h)| + |F(\bar{h})|}$$

$$\text{Fractional change} = \frac{\sum |F(PH)| - |F(P)|}{\sum |F(P)|}$$

where *F*(*P*) and *F*(*PH*) are observed structure factors for native protein and heavy-atom derivatives, respectively. *F_H*(calc) is the r.m.s. value for calculated structure factor contributions of heavy atoms. *E* is the r.m.s. value for the lack-of-closure errors.

apart and are coordinated to main-chain oxygen atoms and to oxygen atoms from the side chains of acidic residues. It is not clear at this stage of the analysis if there are any solvent molecules in the Ca coordination spheres. There is disagreement about the location of high- and low-affinity Ca-binding sites, but most recent studies using Ca binding and NMR seem to agree that sites 3 and 4 are the high-affinity sites²³⁻²⁶. In the Pb derivative, two of the Pb ions replace the Ca ions in domains 1 and 2. Therefore, our data are consistent with other evidence that sites 3 and 4 are the high-affinity sites.

Discussion

Calmodulin exhibits a high degree of sequence homology with parvalbumin and troponin-C^{27,28}, but homology with intestinal Ca-binding protein is restricted to the Ca-binding regions. Based on the structure of parvalbumin, Kretsinger and Barry²⁹ proposed a helix-loop-helix (or EF hand) conformation for the Ca-binding regions in troponin-C. An EF hand is a linear sequence of ~30 amino acids where two nearly perpendicular α -helices flank a 12-residue Ca-binding loop. Parvalbumin contains two such Ca-binding domains. Based on the structural features of these two domains, Tufty and Kretsinger³⁰ postulated that a stretch of 16 amino acids is essential for the formation of a Ca-binding EF hand. They applied this test sequence to a wide variety of proteins and generalized that all intracellular Ca-modulated proteins contain EF hands. The crystal structure determination of intestinal Ca-binding protein¹² and our results for calmodulin provide structural confirmation of this hypothesis. However, there are significant differences among the three structures, primarily in interhelical angles. A comparison of these angles is given in Table 2. The average inter-

Table 2 Inter-helical angles in the Ca-binding domains

Calmodulin		Parvalbumin		Intestinal calcium-binding protein	
Helix pair	Angle (deg)	Helix pair	Angle (deg)	Helix pair	Angle (deg)
I-II	92	C-D	103	I-II	121
III-IV	96	E-F	102	III-IV	121
IV-V	97				
VI-VII	107				

The interhelical angles for parvalbumin and intestinal Ca-binding protein are taken from ref. 12.

helical angle in calmodulin is smaller than in parvalbumin and in intestinal Ca-binding protein. The angles in calmodulin are closest to 90°, whereas those in intestinal Ca-binding protein deviate most from 90°.

Rat testis calmodulin contains two tyrosine residues at positions 99 and 138 that reside in the third and fourth Ca-binding domains, respectively²⁷. Our studies reveal that these two residues occupy different environments. Tyr 99 is directed into the third Ca-binding loop and is fairly accessible to solvent. Tyr 138 points away from loop four into a hydrophobic pocket where it contacts Phe 89 in helix IV and Phe 141 in helix VII. Thus, although Tyr 99 forms contacts only with residues in domain 3, residues from both domains 3 and 4 interact with Tyr 138. The observed structure supports the earlier evidence that Tyr 138 occupies a hydrophobic environment^{8,31-33}. The different environments of the two tyrosine residues could explain why an azido derivative of Tyr 99 can be crosslinked to many more calmodulin-binding proteins than the comparable derivative modified at Tyr 138 (ref. 34). Although Tyr 99 is in one of the Ca-binding loops, it cannot be essential for Ca binding as it is replaced in the spinach protein by phenylalanine and in the protozoan *Tetrahymena* protein by leucine⁵.

Vertebrate calmodulins contain a trimethyllysine residue at position 115. Seamon³¹ has suggested that this amino acid is

exposed to solvent in the absence of Ca. The crystal structure reveals that trimethyllysine is in the region between helices V and VI and is directed towards the solvent with no obvious interactions with other residues in the molecule. Van Eldik *et al.*³⁵ have shown that trimethylation is not required for Ca-dependent interaction with phenothiazines. Putkey *et al.*³⁶ have reported that this post-translation modification also is not required for Ca binding, activation of several calmodulin-dependent enzymes or for Ca-induced conformational changes. These facts, together with the absence of a trimethyl group on Lys 115 in several invertebrate calmodulins^{3,5}, make it unlikely that this rare modification is mandatory for function.

One unusual aspect of the calmodulin structure is the long central helix (helix IV) connecting Ca-binding domains 2 and 3. This long helix seems to serve as the backbone of the protein with the two halves of the molecule attached on either end of the helix. Because this helix is shared by Ca-binding domains 2 and 3, it may have a role in the cooperative binding of Ca. Troponin-C has a similar long central helix that connects the two halves of the molecule; in the crystal structure of troponin-C, there are two Ca ions bound to the carboxy-terminal domain, whereas the cation-binding sites of the amino-terminal domain are Ca free. The amino-terminal portion of troponin-C seems to wrap around the first part of the central helix; it is possible that in a similar fashion the carboxy-terminal portion wraps around the second part of this helix when Ca is removed. Thus, the central helix in troponin-C and calmodulin may be buried in the absence of Ca and exposed when the ion is bound, which would be consistent with proteolysis studies of calmodulin. The greatest sensitivity of calmodulin to trypsin occurs among the residues in this central helix. In the presence of Ca, limited proteolysis results in one cleavage in the centre of this helix that produces two fragments 1-77 and 78-148 (refs 25, 37-39). However, in the absence of Ca the fragments produced are 1-106, 1-90 and 107-148 (refs 25, 37-39). This suggests that Ca binding exposes the residues in helix IV to proteolytic attack.

Several lines of evidence suggest that the residues in helix IV have an important role in the interaction of calmodulin with receptor proteins and drugs. The chemical modification studies of Walsh and Stevens⁴⁰ showed that carboxymethylation of methionine residues 71, 72 and 76 prevents the productive interaction of calmodulin with cyclic phosphodiesterase. Also, Newton *et al.*⁴¹ have covalently attached a phenothiazine to the Ca²⁺/calmodulin complex. The primary site of attachment is Lys 75, which resides in helix IV (D. L. Newton and C. B. Klee, personal communication). Crystallographic studies of calmodulin complexes with drugs and receptor proteins will be of particular interest in elucidating the role of helix IV in these interactions.

One approach that can now be used to study structure/function relationships for calmodulin is to alter appropriate regions of the molecule to see how these modifications affect enzyme binding and activation. This is feasible as the chicken calmodulin gene has been cloned^{6,15} and the cDNA has been introduced into bacterial expression vectors to yield large amounts of easily isolatable protein³⁶. The chicken calmodulin gene is composed of eight exons and seven introns; five of the introns are in the amino-acid coding region¹⁵. One intron separates the DNA that encodes amino acids 82 and 83, thus interrupting helix IV and serving to separate the two homologous halves of the protein. Of the other helices, introns split only two: helix I between amino acids 10-11, a region not involved in Ca binding, and helix VII, between amino acids 139-140. This latter intron interrupts Ca-binding site 4. The remaining two introns interrupt Ca-binding loops 2 and 3, respectively. These data reveal that three of the four Ca-binding domains are interrupted by introns and therefore do not totally support the proposal that introns primarily divide the functional domains of proteins^{42,43}. The current theory suggests that introns separate regions of proteins that are on the surface of the molecule⁴⁴. The locations of all the introns in the calmodulin gene support this latter theory.

We thank Charles R. Mena and James G. Chafouleas for

assistance in the preparation of calmodulin samples. This research was supported by NIH grants GM-29427, CA-13148, DE-02670 and HD-07503. W.J.C. is the recipient of NIH

Research Career Development Award CA-00696. J.S.S. is the recipient of a postdoctoral fellowship under NIH Training Grant CA-09128.

Received 8 February; accepted 14 March 1985.

1. Baba, M. L., Goodman, M., Berger-Cohn, J., DeMaille, J. G. & Matsuda, G. *Molec. Biol. Evol.* **1**, 442-455 (1984).
2. Means, A. R., Tash, J. S. & Chafouleas, J. G. *Physiol. Rev.* **62**, 1-39 (1982).
3. Klee, C. B. & Vanaman, T. C. *Adv. Protein Chem.* **35**, 213-321 (1982).
4. Chafouleas, J. G., Dedman, J. R., Munjaal, R. P. & Means, A. R. *J. Biol. Chem.* **254**, 10262-10267 (1979).
5. Marshak, D. R., Clarke, M., Roberts, D. M. & Watterson, D. M. *Biochemistry* **23**, 2891-2899 (1984).
6. Putkey, J. A. *et al. J. Biol. Chem.* **258**, 11864-11870 (1983).
7. Dedman, J. R., Potter, J. D., Jackson, R. L., Johnson, J. D. & Means, A. R. *J. Biol. Chem.* **252**, 8415-8422 (1977).
8. Richman, P. G. & Klee, C. B. *Biochemistry* **17**, 928-935 (1978).
9. Kilhoffer, M. C., Demaille, J. G. & Gerard, D. *Biochemistry* **20**, 4407-4414 (1981).
10. Anderson, J. M., Charbonneau, H. J., Jones, H. P., McCann, R. O. & Cormier, M. J. *Biochemistry* **19**, 3113-3120 (1980).
11. Kretsinger, R. H. & Nockolds, C. E. *J. Biol. Chem.* **248**, 3313-3326 (1973).
12. Szebenyi, D. M. E., Obendorf, S. K. & Moffat, K. *Nature* **294**, 327-332 (1981).
13. Herzberg, O. & James, M. N. G. *Nature* **313**, 653-659 (1985).
14. Sundaralingam, M. *et al. Science* **227**, 945-948 (1985).
15. Simmen, R. C. M. *et al. J. Biol. Chem.* **260**, 907-912 (1985).
16. Cook, W. J. & Sack, J. S. *Meth. Enzym.* **102**, 143-147 (1983).
17. North, A. C. T., Phillips, D. C. & Mathews, F. S. *Acta crystallogr.* **A24**, 351-359 (1968).
18. Blow, D. M. & Crick, F. H. C. *Acta crystallogr.* **12**, 794-802 (1959).
19. Terwilliger, T. C. & Eisenberg, D. *Acta crystallogr.* **A39**, 813-817 (1983).
20. Rossmann, M. G. *Acta crystallogr.* **13**, 221-226 (1960).
21. Bricogne, G. *Acta crystallogr.* **A32**, 832-847 (1976).

22. Jones, T. A. *J. appl. Crystallogr.* **11**, 268-272 (1979).
23. Klee, C. B. *Biochemistry* **16**, 1017-1024 (1977).
24. Haiech, J., Klee, C. B. & Demaille, J. G. *Biochemistry* **20**, 3890-3897 (1981).
25. Andersson, A., Forsen, S., Thulin, E. & Vogel, H. J. *Biochemistry* **22**, 2309-2313 (1983).
26. Klevit, R. E., Dalgarno, D. C., Levine, B. A. & Williams, R. J. P. *Eur. J. Biochem.* **139**, 109-114 (1984).
27. Dedman, J. R., Jackson, R. L., Schreiber, W. E. & Means, A. R. *J. Biol. Chem.* **253**, 343-346 (1978).
28. Watterson, D. M., Sharief, F. & Vanaman, R. C. *J. Biol. Chem.* **255**, 962-975 (1980).
29. Kretsinger, R. H. & Barry, C. D. *Biochim. biophys. Acta* **405**, 40-52 (1975).
30. Tufty, R. M. & Kretsinger, R. H. *Science* **187**, 167-169 (1975).
31. Seamon, K. B. *Biochemistry* **19**, 207-215 (1980).
32. Hincke, M. T., Sykes, B. D. & Kay, C. M. *Biochemistry* **20**, 4185-4193 (1981).
33. Ikura, M. *et al. Biochemistry* **22**, 2573-2579 (1983).
34. Klevit, R. E. & Vanaman, T. C. *J. Biol. Chem.* **259**, 15414-15424 (1984).
35. Van Eldik, L. J., Grossman, A. R., Iverson, D. B. & Watterson, D. M. *Proc. natn. Acad. Sci. U.S.A.* **77**, 1912-1916 (1980).
36. Putkey, J. A., Slaughter, G. R. & Means, A. R. *J. Biol. Chem.* (in the press).
37. Drabikowski, W., Kuznicki, J. & Grabarek, Z. *Biochim. biophys. Acta* **485**, 124-133 (1977).
38. Newton, D. L., Oldewurtel, M. D., Krinks, M. H., Shiloach, J. & Klee, C. B. *J. Biol. Chem.* **259**, 4419-4426 (1984).
39. Guerin, D., Krebs, J. & Carafoli, E. *J. Biol. Chem.* **259**, 15172-15177 (1984).
40. Walsh, M. & Stevens, F. C. *Biochemistry* **17**, 3924-3928 (1978).
41. Newton, D. L., Burke, T. R. Jr, Rice, K. C. & Klee, C. B. *Biochemistry* **22**, 5472-5476 (1983).
42. Gilbert, W. *Nature* **271**, 501 (1978).
43. Stein, J. P., Catterall, J. F., Kristo, P., Means, A. R. & O'Malley, B. W. *Cell* **21**, 681-687 (1980).
44. Lonberg, N. & Gilbert, W. *Cell* **40**, 81-90 (1985).

LETTERS TO NATURE

Low frequency variability and interstellar focusing

T. V. Cawthorne & B. J. Rickett*

Cavendish Laboratory, Madingley Road, Cambridge CB3 0HE, UK

The flux variations of extragalactic radio sources at low radio frequencies have proved very difficult to explain as changes in the luminosity of a simple synchrotron source. To circumvent this problem, several more complicated source models, such as those involving relativistic expansion of the source or coherent radiation mechanisms, have been devised. However, it has been suggested recently that the observed flux changes may be the result of a focusing effect of irregularities in the local interstellar medium. Here we present evidence which suggests that, in accordance with this explanation, low frequency variability is a more common feature of sources at low galactic latitudes. In the sample of radio sources monitored over a period of five years at Bologna, the mean galactic latitude of the variable sources is smaller than that of the non-variables to an extent that would occur with a chance of only 2% were the two groups drawn from the same parent distribution.

It is now firmly established that the low frequency (≤ 1 GHz) flux density of some extragalactic radio sources varies on timescales of several months: the first conclusive evidence was provided by Hunstead¹ and there have since been several major monitoring programmes to search for variability—notably, those of Fanti *et al.*^{2,3}, using the Bologna telescope at 408 MHz, and Spangler and Cotton⁴. There are now more than 50 known low-frequency variable sources; most have flat radio spectra, normally associated with compact sources, and many are variable at higher frequencies.

If the flux variations are caused by changes in the source luminosity, the timescale on which they occur may be used to

obtain an upper limit for the size of the source. Because the disturbance responsible for the change propagates with a speed less than that of light (c), the source diameter must be less than c (variation timescale)/(1+z) where z is the redshift. However, variations in amplitude of several janskys on timescales of months in sources of moderate or high redshift require very large radiation brightness temperatures, well in excess of those that can be obtained from an incoherent synchrotron source⁵. Recently, however, Rickett *et al.*⁶ have suggested that the variations are, at least in part, not intrinsic to the source, but the result of a refractive focusing effect of the general interstellar medium. (A similar suggestion was made earlier by Shapirovskaya⁷, in which the focusing irregularities were concentrated in galactic loop structures.)

Small-scale irregularities in the density of interstellar electrons scatter waves into an angular spectrum of radius θ_{sc} and, for long wavelengths λ (typically ≥ 10 cm), interference among its components causes 100% modulations in the intensity of a point source, on scales of $\lambda/(2\pi\theta_{sc})$ in the plane normal to the direction of the source. It has been noted⁸ that these modulations account for the scintillation of pulsars observed on timescales of order 10 min, at frequencies below 1 GHz. Assuming a power law spectrum of irregularities with power per unit wavenumber proportional to (wavenumber)^{-(2+ α)}, and a Kolmogorov spectrum ($\alpha = 5/3$), recent studies of pulsar scintillation⁹ allow, that for a path length L through the galaxy, $\theta_{sc} = 6(\lambda \text{ m}^{-1})^{11/5}(L \text{ kpc}^{-1})^{3/5}$ marc s, which applies for a plane-wave source. The distribution of scattering irregularities is often approximated by a layer of thickness ± 500 pc centred on the galactic plane, so that at galactic latitude b , $(L \text{ kpc}^{-1}) = 0.5 \text{ cosec}(b)$ and $\theta_{sc} = 4(\lambda \text{ m}^{-1})^{11/5}(\text{cosec } b)^{3/5}$ marc s. These diffractive scintillations have never been observed for extragalactic radio sources^{10,11}, which is usually taken to mean that the angular diameters of such sources, θ_0 , are sufficiently large that interference patterns from adjacent parts of the source are smeared over a distance $L\theta_0$, much larger than the scale of the patterns themselves, $\lambda/(2\pi\theta_{sc})$. The intensity modulations are thus completely levelled. However, the large-scale irregularities present can also modulate the intensity by refractive focusing, which produces caustics in the wavefront from the source that are noticed as intensity fluctuations as the observer passes through them. The largest scale in the medium that

* Permanent address: Department of Electrical Engineering and Computer Science, University of California, San Diego, La Jolla, California 92043, USA.

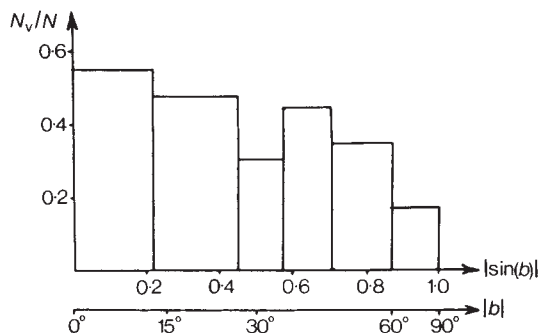


Fig. 1 The height of each bin represents the fraction (N_v/N) of sources from the Bologna sample that show fluctuations of amplitude $>5\%$ of the mean. In order of increasing $|b|$, the values for (N_v/N) are: 11/20, 9/19, 5/17, 9/20, 7/20 and 3/18.

can influence the radiation at any one point, is determined by the effective scattering disk of radius $L\theta_{sc}$ at distance L . For a power-law wavenumber spectrum, the increasing power in larger scales results in substantial modulation in the intensity of a point source on scales of $L\theta_{sc}$. The exact level of modulation depends on the power law exponent α and is particularly strong for exponents above four. Nevertheless, measurements of laser scintillation in the atmosphere show that even Kolmogorov spectra give rise to 50–100% modulation¹². Similar results are found for pulsars in the interstellar medium⁶. Compact extragalactic sources of diameter θ_0 should fluctuate as do pulsars if they are small enough ($L\theta_0 < L\theta_{sc}$). Even though this limit is much less stringent than that for diffractive scintillation, low frequency VLBI observations¹³ suggest that, at metre wavelengths, most compact sources have $\theta_0 \geq \theta_{sc}$, so the refractive scintillations are at least partially quenched by smearing over a scale $L\theta_0$. The resulting modulation will be $\Delta S_{rms}/S \approx \theta_{sc}/\theta_0$ on timescale $L\theta_0/2v$ where v is the relative velocity of the Earth and the scintillation pattern (10 – 100 km s^{-1}).

The foregoing theory makes predictions for the dependence of source variability on galactic latitude, which we have examined using the Bologna sample of Fanti *et al.*³. Figure 1 shows the most interesting result; the fraction of sources monitored that fluctuate with amplitude $>5\%$ of the mean is shown for each of six galactic latitude bins as a function of $|\sin(b)|$. The definite fall in this fraction with increasing latitude suggests that the variations are related to propagation in the interstellar medium. Because a χ^2 test takes no account of the order of the bins, it is not expected to attach great significance to this trend and, indeed, it does not. The Mann–Whitney test¹⁴, which relies on the difference between the ranks of the variable and non-variable sources with galactic latitude, is far more appropriate. This test confirms that the variable sources have statistically lower galactic latitudes than those with steady fluxes; it yields a 2% chance that the two populations would be drawn in the manner observed from the same distribution in b . However, the possibility that this trend is caused by selection effects must be examined.

The Bologna sample is made up of several subsamples, consisting of sources chosen by known variability, flat spectra or interplanetary scintillations. To produce the observed trend in a set of intrinsically variable sources by selection, at least one subsample would have to be strongly biased, either in favour of low latitudes, with an excess of variables, or in favour of high latitudes, with a deficiency of variables. Such an effect might arise if one subsample, biased towards low or high latitudes, was selected at very high or very low frequencies, respectively. One would then expect either an excess of flat spectrum sources (likely to be variable) at low galactic latitudes, or steep spectrum and probably extended sources (unlikely to vary) at high latitudes. However spectral information exists for most of the sources and spectral indices $-\log(\text{flux})/\log(\text{frequency})$

can be found. Taking a flat spectrum source to have spectral index <0.5 in the region 0.5–5.0 GHz, there is no evidence to suggest that the flat and steep spectrum sources in the Bologna sample are differently distributed in b .

One may also restrict attention to sources in a narrow band of spectral index, across which there is little change in likelihood of variability. Almost all the sources with spectral index <0.2 are variable, but in the range of spectral indices 0.2–0.5, where there is no significant trend in spectral index with $|b|$, the variable sources again lie at systematically lower galactic latitudes than the non-variables, with a chance of $<1\%$ that the two populations would be drawn from the same distribution in b . Furthermore, the significance of the trend depends strongly on the dearth of variable sources above $|b|=60^\circ$, which is apparent (although not highly significant) in the larger subsamples, individually. Therefore, it does not seem likely that the observed trend was produced by a dependence of spectral index on $|b|$, such as may arise through compiling the whole Bologna sample from its components. We cannot, of course, eliminate the possibility that unknown systematic effects are responsible for the presence of the trend within the subsamples.

Although the exact form that the focusing mechanism predicts for Fig. 1 depends on the source model assumed and the (unknown) distribution of angular sizes on milliarcsecond scales, the observed trend is clearly in the right sense. At a given latitude, the fraction of uniform sources that fluctuate with amplitude $>5\%$ of the mean is, from the above expression for $\Delta S_{rms}/S$, just that with angular size $\theta_0 < 20\theta_{sc}$, which will increase with θ_{sc} and hence with $\text{cosec}(b)$. However, the sources will probably be inhomogeneous and, if at some $|b|$ θ_{sc} becomes sufficiently large that the scintillating components of most compact sources fluctuate with amplitude $>5\%$ of the total flux, then one would expect no further increase in the fraction of variable sources at lower galactic latitudes. This may account for the absence of a marked increase in the probability of variability at low b , where θ_{sc} increases most dramatically.

Using published data for the variable sources in the Bologna sample¹⁵, we have looked for relations between galactic latitude and the amplitude and timescales of variation. None of the expected trends are apparent in these data, which may be because the published parameters describe individual outbursts rather than r.m.s. variations. However, wide ranges in the fraction of flux coming from the compact component, in angular size θ_0 and in velocity of the Earth through the scintillation pattern, could also mask these effects. When metre wavelength VLBI data become available, the flux and diameter of a compact component may be used to explore further the relationships between amplitude, timescale, diameter and galactic latitude.

Although Fig. 1 suggests that the focusing mechanism is responsible for low frequency variations in many sources, there remains the possibility that some variations are intrinsic. In this context we note that, of six sources recently monitored¹⁶, five varied almost simultaneously over a range of frequencies between 300 and 1,400 MHz, as the focusing mechanism predicts. One, however, exhibited brightening, with delays at lower frequencies, of the kind often associated with outbursts in a synchrotron source; this source may well be intrinsically variable.

Any source of observed diameter of the order of 10 mas at a few hundred MHz that shows angular broadening, diminishing in diameter at higher frequencies, has $\theta_0 \leq \theta_{sc}$, and should undergo refractive scintillations with 50–100% modulation as do pulsars, on timescales comparable with other extragalactic sources. Such sources are important candidates for monitoring. It has been noted¹⁷ that at very low galactic latitudes ($\leq 5^\circ$), particularly towards the galactic centre, broadening is rather greater than is expected by extrapolating from that at high latitudes using the $\text{cosec}(b)$ relation for L . Small-diameter sources seen along such paths should also fluctuate at centimetre wavelengths and it is plausible that several of the variable low-latitude sources, reported by Ryle *et al.*¹⁸ and by Taylor

and Gregory¹⁹, are extragalactic sources whose flux is modulated by refractive scintillation.

We thank Dr. L. Padrielli for discussion on the Bologna sample, Stephen Eales for advice on matters statistical, and Drs A. J. B. Downes and P. A. G. Scheuer for helpful comments.

Received 19 November 1984; accepted 28 February 1985.

- Hunstead, R. W. *Astrophys. Lett.* **12**, 193–200 (1972).
- Fanti, R., Ficarra, A., Mantovani, F., Padrielli, L. & Weiler, K. *Astr. Astrophys. Suppl. Ser.* **36**, 359–369 (1979).
- Fanti, C. et al. *Astr. Astrophys. Suppl. Ser.* **45**, 61–78 (1981).
- Spangler, S. R. & Cotton, W. D. *Astr. J.* **86**, 730–746 (1981).
- Hoyle, F., Burbidge, G. & Sargent, W. L. W. *Nature* **209**, 751–753 (1966).
- Rickett, B. J., Coles, Wm. A. & Bourgois, G. *Astr. Astrophys.* **134**, 390–395 (1984).
- Shapirovskaia, N. Ya. *Sov. Astr.* **22**, 544–547 (1978).
- Scheuer, P. A. G. *Nature* **211**, 920–922 (1968).
- Cordes, J. M., Weisberg, J. & Boriakoff, V. *Astrophys. J.* (in the press).
- Lovelace, R. V. E. & Backer, D. C. *Astrophys. Lett.* **11**, 135–138 (1972).
- Dennison, B. & Condon, J. J. *Astrophys. J.* **246**, 91–99 (1981).
- Coles, Wm. A. & Frehlich, R. G. *J. opt. Soc. Am.* **72**, 1042–1048 (1982).
- Wilkinson, P. N., Readhead, A. C. S., Anderson, B. & Purcell, G. H. *Astrophys. J.* **232**, 365–381 (1979).
- Conover, W. J. *Practical Nonparametric Statistics* (Wiley, New York, 1980).
- Fanti, C. et al. *Astr. Astrophys.* **118**, 171–179 (1983).
- Dennison, B. et al. *Astrophys. J.* **281**, L55–L59 (1984).
- Cordes, J. M., Ananthakrishnan, S. & Dennison, B. *Nature* **309**, 689–691 (1984).
- Ryle, M., Caswell, J. L., Hine, G. & Shakeshaft, J. *Nature* **276**, 571–573 (1978).
- Taylor, A. R. & Gregory, P. C. *Astr. J.* **88**, 1784–1809 (1983).

Fermionic and bosonic interference during spin rotation through 2π radians

Herbert J. Bernstein

Massachusetts Institute of Technology, Cambridge,
Massachusetts 02139
and *Hampshire College, Amherst, Massachusetts 01002, USA

Spinor wavefunctions change their sign under 2π rotation, an effect observable in interference experiments^{1,2}. Several experiments have used magnetic precession of neutrons to implement rotations^{3–7}. Other investigators have observed phenomena in molecular, nuclear or atomic systems with integral spin that have been identified as interference effects produced by 2π rotation^{8–11}. I describe here a distinction between experiments on spin-1/2 neutrons and those on bosons, one that is crucial to interpreting these experimental results. Following recent work on interpreting magnetic precession as rotation^{12,13}, I show how to resolve Zeilinger's thought experiment¹², which actually measures the spin direction of two partial beams, with Byrne's point¹³ that the recombined beam of neutrons has been rotated only π rad when the relative rotation between two partial beams is 2π . The essential distinguishing step lies in recognizing the lack of geometric significance, in general, of the mathematical parameter that equals 2π in the boson case.

When one transforms spin-1 states by true spatial rotation, the smallest covering angle (the first angle that restores original behaviour to a continuously rotated two-state subsystem) is π rad, not 2π rad. The observed sign change is nothing more than the (-1) of complete mirror reflection; such a reflection within a plane is exactly equivalent to a π rotation in three-dimensional space. This touches on the central problem of interpretation 'Is there any dynamic interaction that we may interpret as a pure rotation?'¹⁴.

Although static magnetic fields generally cause additional effects, a uniform magnetic field that turns on after the particle enters does implement a pure spin rotation. Here I suggest a new symmetrical neutron interference experiment, which, with static fields, clarifies why the spin of the recombined beam has rotated only π rad; with time-dependent fields, this configuration provides a feasible implementation of pure rotation. Indeed, the question posed above is addressed in the context of broader issues in the interpretation of quantum theory at the

end of this letter. I begin by considering the relationship between rotations and dynamic interactions.

In quantum theory, certain unitary transformations represent spatial rotations, whereas others are used to represent the time evolution of a system. A rotation through angle θ in the x direction, for example, is $R_x(\theta) = \exp(-i\theta J_x/\hbar)$, where J_x is the operator for the x component of angular momentum, which may be calculated from the derivative, or generator, of the rotation

$$J_x/\hbar = i dR_x/d\theta|_{\theta=0} \quad (1)$$

Similarly, the hamiltonian, H , is a generator of the unitary time evolution operator, $U(t)$. If $|\Psi(t)\rangle = U(t)|\Psi(0)\rangle$, then the Schroedinger equation states

$$H/\hbar = i dU/dt|_{t=0} \quad (2)$$

assuming that H is time independent. The complications introduced to equation (2) by time-dependent interactions are exactly analogous to finding the angular momentum as a derivative at $\theta \neq 0$.

The analogy between equations (1) and (2) underlies all of the work on boson systems. In the simplest case, one finds a two-state subsystem where the hamiltonian induces transformations from one state to the other and vice versa, with equal transition probability, A , given by

$$H/\hbar = \begin{pmatrix} 0 & A \\ A & 0 \end{pmatrix} \quad (3)$$

which has the same form as the Pauli matrix for spin $\frac{1}{2}$,

$$\sigma_x/2 = J_x/\hbar = \begin{pmatrix} 0 & \frac{1}{2} \\ \frac{1}{2} & 0 \end{pmatrix} \quad (4)$$

Indeed, just as the general rotation

$$R_x^{1/2}(\theta) = \begin{pmatrix} \cos(\theta/2) & -i \sin(\theta/2) \\ -i \sin(\theta/2) & \cos(\theta/2) \end{pmatrix} \quad (5)$$

reduces to a phase factor (-1) when $\theta = 2\pi$, so a suitable evolution time for the hamiltonian in equation (3) also produces such a sign change. It is essentially this sign change that is measured by all boson experiments. The analogy between time evolution and rotation is the essence of the work on integral spin systems.

On the other hand, the idea of the neutron precession experiment relies on identifying the hamiltonian with the angular momentum. Indeed, to produce a true rotation, the hamiltonian must be proportional to the actual angular momentum of the system and not just to an analogue of a spin- $\frac{1}{2}$ operator^{2,14}. This proportionality to angular momentum has double significance for interpreting the experiments: (1) the correspondence principle teaches that when energy is proportional to angular momentum, the interaction exerts a torque on the system, thereby causing it to rotate; (2) the parameter in equation (5) corresponds (as in the *gedanken* Stern-Gerlach experiments of ref. 12, or in its more feasible experiments using polarized neutron birefringence) to a measurable angle. By contrast, an actual 2π rotation of the analogous boson two-state subsystem produces no change of sign; the measurable rotation angle that results in sign change is π rad instead. This can be seen by considering the spin-1 example.

For a system of spin 1, with the usual basis set $\{|+1\rangle, |0\rangle, |-1\rangle\}$, labelled by $J_z/\hbar$, the generator of a z -axis rotation is diagonal. But transforming to a basis of $|x\rangle = 1/\sqrt{2}(|+1\rangle + |-1\rangle)$, $|y\rangle = 1/\sqrt{2}(|+1\rangle - |-1\rangle)$ and $|z\rangle = |0\rangle$, we find

$$J_z/\hbar = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad (6)$$

which clearly has a two-dimensional subspace (spanned by $\{|x\rangle, |y\rangle\}$) where it takes the off-diagonal form required by equation (3). Unlike the spin- $\frac{1}{2}$ generator (equation (4)),

* Address for correspondence.

however, the non-zero elements are 1, not $\frac{1}{2}$. This difference makes the physical rotation, restricted to the $\{|x\rangle, |y\rangle\}$ subspace,

$$R_z^{1, \text{restricted}}(\theta) = \begin{pmatrix} \cos \theta & -i \sin \theta \\ -i \sin \theta & \cos \theta \end{pmatrix} \quad (7)$$

in clear distinction to the half-angle dependence in equation (5). Here it is obvious that the physical π rotation reduces to a sign change. Indeed, just as the states $|+1\rangle$ and $|-1\rangle$ correspond to spin fully up and down along the z axis, so the states $|x\rangle$ and $|y\rangle$ represent spin along the x and y axes, respectively. These act like well-behaved vectors in all respects, including complete return (no sign change) under 2π rotation and inversion when rotated by π rad about an axis perpendicular to their plane. Note that under $\pi/2$ rotation $|x\rangle$ becomes $-i|y\rangle$ and $|y\rangle$ becomes $-i|x\rangle$.

This last fact is probably the source of confusion in over-emphasizing the spinor analogy, for it is a π rotation that converts a spinor into its orthogonal complement. If the experiment (or, more exactly, its interpretation) identifies this conversion of $|x\rangle$ into $|y\rangle$ and vice versa with a spinorial π rotation, the analogy of doubling the physical angle is established. As a result, the first return to the initial states is analogous to a spinorial 2π rotation. Equation (7) shows this 'covering' return is actually a π rotation in physical space.

For non-scalar particles travelling at the speed of light, only two angular momentum states exist, which, in the case of photons, are the two distinct linear polarizations (and/or combinations thereof). If the direction of travel is the z axis, then the polarization of a single photon is a complex unit two vector in the x - y plane, essentially the direction and phase of the electric field vector of the corresponding electromagnetic wave. The physical rotation through angle ϕ about $\hat{z}$ (the caret denotes a real unit vector) is

$$R_z(\phi) = \begin{pmatrix} \cos \phi & -i \sin \phi \\ -i \sin \phi & \cos \phi \end{pmatrix}$$

in a basis where

$$\hat{x} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \hat{y} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

Once again we see the full-angle dependence.

The space of all distinct polarization vectors, $\begin{pmatrix} a \\ b \end{pmatrix}$, with $|a|^2 + |b|^2 = 1$, can be parametrized by points on a unit sphere in a fictitious three-dimensional space, where rotations are spinorial by definition. One can introduce angles ξ, η , exactly analogous to the usual θ, ϕ angles on a sphere, but the north and south poles of the ξ, η sphere are the spinor basis states, $\hat{x}$ and $i\hat{y}$. Orthogonal directions in space are thus represented by polar opposite points in spinor (ξ, η) space and the spinorial angle between states is twice the angle in physical space. Again, the 2π rotation in a fictitious spinor space is actually a π rotation in real space. This fictitious ('Faraday') space, with its rotation analogies, is of great use in understanding and picturing the relations among states of polarization, but its transformations are not physical space rotations¹⁵.

The distinction between fermions and bosons can be extended to higher spins. For all fermions, true 2π rotations produce a sign change, whereas for bosons they do not. The smallest rotation angle that produces a sign change in a two-state subspace of spin J is π/J . For fermion spin greater than $\frac{1}{2}$ this is an odd submultiple of 2π , which means that a 2π rotation may be composed of these smaller rotations, by multiplying them together. Then the (-1) factor appears raised to an odd-integer power; the full 2π rotation, therefore, also changes sign. For bosons π/J is an integral submultiple of π and a 2π rotation therefore produces the factor $(-1)^{2J}$, that is, no sign change.

In general, taking m (the angular momentum component along the rotation axis) as a label, we have $R(\theta)|m\rangle = \exp(-im\theta)|m\rangle$. The value of m ranges from $-J$ to $+J$ in steps of one unit, where J is the total spin (half-integral for fermions, integral for bosons). Because this includes the values $m = \pm 1$

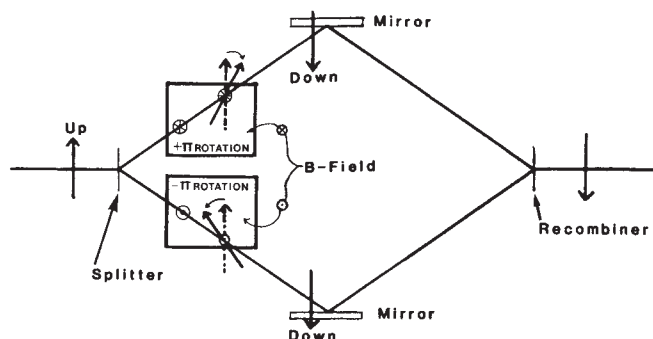


Fig. 1 A symmetric magnetic precession experiment (schematic).

for non-scalar bosons and $m = \pm \frac{1}{2}$ for fermions, the smallest angle producing a sign change of the entire space of states is always π for bosons and 2π for fermions.

In the highest and lowest m ($= \pm J$), choosing the basis $|+\rangle = 1/\sqrt{2}(|J\rangle + |-J\rangle)$ and $|-\rangle = 1/\sqrt{2}(|J\rangle - |-J\rangle)$ makes the rotation matrix

$$R_z^{J, \text{restricted}}(\theta) = \begin{pmatrix} \cos(J\theta) & -i \sin(J\theta) \\ -i \sin(J\theta) & \cos(J\theta) \end{pmatrix} \quad (8)$$

from which it follows that sign change first occurs for $\theta = \pi/J$, as stated above. This is particularly important for the case of a particle travelling at the speed of light, because its two polarization states are precisely $|+J\rangle$ and $|-J\rangle$, hence $|+\rangle$ and $|-\rangle$ are a natural basis for the rotations. The cases of $J = \frac{1}{2}$ and 1, corresponding to neutrino and photon respectively, fit formulas given above. The one other massless entity in modern physics, that is the 'graviton', has spin 2. Pictures (p. 230 of ref. 16) of acceleration for linearly polarized gravitational waves clearly show the rotation effect described above: the value of J for bosons denotes a J -fold symmetry of the transverse oscillations in the case of light-like radiation.

What of the interpretation of the neutron spin precession as rotation? Zeilinger's clever thought experiment¹² actually measures the spin direction of two partial beams, providing a physical demonstration of the rotation, whereas Byrne¹³ points out that the recombined beam has been rotated by only π rad when the relative rotation between partial beams is 2π .

That both authors are correct is directly demonstrated by a 2π -rotation experiment performed symmetrically, with a positive π rotation along one path and a negative (or opposite-sense) π rotation along the other path. In some cases, this symmetrical rotation scheme could help eliminate possibilities for determining along which path the neutron travelled, for example, by using a common source for the two fields and arranging that any magnetic energy exchange be the same for both possible paths. As is well known in quantum theory, if one can tell which path was taken, the phase coherence between the paths is destroyed and the interference is not observable. If the 2π rotation experiment were done asymmetrically, the passage of a neutron dipole moment through the magnetic field along one path might conceivably produce detectable back reaction, signalling which path it took. Almost certainly there would be some additional phase shift, beyond the expected sign change resulting from 2π rotation, because of the asymmetry.

If we examine the symmetrical experiment (see Fig. 1), we immediately notice that the π rotations produce the same orientation of neutron spin, whichever path is taken. Thus, if the initial spin is 'up' (as in Fig. 1), then the spin of each partial beam will be 'down' when they are about to recombine. The recombined beam is also spin 'down'. Note that this experiment would not require a polarized initial beam because the same π rotation effect occurs with a spin 'down' particle. Of course, to follow the spin by, say, Zeilinger's Stern-Gerlach apparatus would require a polarized beam.

These remarks about orientation apply equally to bosons. For example, consider a spin-1 case and let 'up' denote the state $|x\rangle$

defined above. Equation (7), with $\theta = \pi$, shows that positive π rotation transforms $|x\rangle$ to $-|x\rangle$. But so does $\theta = -\pi$; the recombined beam points in the 'down' direction, opposite 'up'. This accords with Fig. 1, but produces no relative phase between the two beams, so there is no difference in the interference pattern from the case of zero rotation. An observable sign change occurs at smaller rotation, namely, $\theta = \pm\pi/2$.

The fact that the combined spin direction after a 2π relative rotation is only π rad away from the initial direction is a special case of superposition of two equal-intensity coherent spinors. In general, the spin of the superposition lies in the plane bisecting the angle between the two spin vectors. This can be seen quickly in spinor algebra. For simplicity, take the x - z plane to contain both the combining spinors and let the x axis be their angle bisector (see Fig. 2). This is just a choice of coordinate axes. If we use the convention that

$$|\theta, \phi\rangle = \begin{pmatrix} \cos(\theta/2) & e^{-i\phi/2} \\ \sin(\theta/2) & e^{+i\phi/2} \end{pmatrix} \quad (9)$$

where θ, ϕ are the colatitude and azimuthal angles, respectively, we have

$$\begin{aligned} |\psi^{\text{recombined}}\rangle &\propto \left| \frac{\pi}{2} - \eta, 0 \right\rangle + \left| \frac{\pi}{2} + \eta, 0 \right\rangle \\ |\psi^{\text{recombined}}\rangle &\propto \cos \frac{\eta}{2} \begin{pmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{pmatrix} = \cos \frac{\eta}{2} \left| \frac{\pi}{2}, 0 \right\rangle \end{aligned} \quad (10)$$

If we take η as the individual rotation angle, π , in the symmetric 2π rotation experiment, then equation (10) implies a vanishing amplitude for the recombined state that reflects the perfect destructive interference attributable to the (-1) phase factor induced by one full rotation. (In the notation of Byrne's paper¹³, $\beta = 2\pi$; and $A = B$, that is the parameters characterizing the intensity of the recombining beams are equal, imply the vanishing of the recombined beam itself, for no particles (ideally) would be transmitted. Byrne's equation (3) reflects this—the transmission coefficient drops to zero for these values. His point about the spin polarization is nevertheless correct; it does suffer a rotation of π rad. Experimental confirmation would require taking the limits as $\beta \rightarrow 2\pi$ and seeing the polarization approximate rotation by π as closely as one likes.)

It is interesting briefly to consider the effects of a relative phase shift, α , on the recombined beams. In this case, spinor addition gives

$$\begin{aligned} |\phi^{\text{recombined}}\rangle &\propto \left(e^{-i\alpha/2} \left| \frac{\pi}{2} - \eta, 0 \right\rangle + e^{+i\alpha/2} \left| \frac{\pi}{2} + \eta, 0 \right\rangle \right) \\ &= \sqrt{\cos^2(\alpha/2) \sin^2(\eta/2) + \sin^2(\alpha/2) \cos^2(\eta/2)} \\ &\quad \times \left| \frac{\pi}{2}, 2\tan^{-1}(\tan(\eta/2) \tan(\alpha/2)) \right\rangle \end{aligned} \quad (11)$$

Previous work^{2,14} has discussed the fact that a static magnetic field introduces precisely such a relative phase shift, as well as the desired rotation, in an asymmetrical relative rotation experiment. This small extra phase shift arises because the static configuration conserves energy. Neutrons with spin parallel to the field increase their kinetic energy, spending less time in the field than those aligned oppositely. The resulting phase shifts are not precisely equal and opposite. That in such an experiment the rotation is not pure, would be revealed in these spin-tracking experiments by arriving at a recombined polarization that lies in the x - y plane, but not precisely along the x axis.

Note that various interactions can produce the hamiltonian of equation (3), which results in a time evolution operator

$$U(t) = \begin{pmatrix} \cos(At) & -i \sin(At) \\ -i \sin(At) & \cos(At) \end{pmatrix} \quad (12)$$

For the slightly more complicated case where A is a function

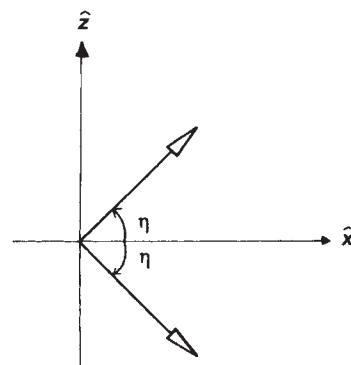


Fig. 2 Addition of spinors.

of time, the product At is replaced by the integral $I(t) = \int_0^t A(\tau) d\tau$. Then

$$U(t) = \begin{pmatrix} \cos I(t) & -i \sin I(t) \\ -i \sin I(t) & \cos I(t) \end{pmatrix}$$

But the relationship between the product At , or $I(t)$, and a physical rotation angle (if any) depends on the angular momentum, J , of the system. H must be proportional to J to identify $U(t)$ as a rotation. The advantage of magnetic precession as a means of performing rotation lies in the well-established proportionality of the hamiltonian to the magnetic moment and hence to angular momentum. Even so, as stated above (and developed previously^{12,14,17}) certain magnetic field distributions produce additional phase shifts. Thus, one must be careful of the theoretical analysis that undertakes identification of any dynamical process with a rotation. For it is well known that interactions produce phase shifts that can mimic those produced by physical rotations of either the reference system (passive rotations) or of the state-preparation devices (active rotations).

More directly, to identify a transformation as a rotation requires a physical rotation angle and preferably one that is, in principle, directly measurable. Zeilinger's analysis shows how to make such measurements during magnetic precession. The issue of such a geometric (or macroscopic) correlative of the spin, for polarized neutrons, undergoing precession has been raised in the earlier literature on the interpretation of these experiments^{13,17-21}. The suggestion of using a strong static helical field (whose change of direction in space could be easily measured) was an early step in this direction²¹. Such a field entrains the moment of the polarized neutron and guides it around in space. The theory shows that in the asymmetric case such a helical static configuration also produces only approximate rotations^{14,20-22}. The approximation involved is so good (often better than 1 part per 10^{12}) that it enters the discussion only at the level of interpretation. Nevertheless, every time-independent magnetic field thus far examined requires these approximations in order to be interpreted as a true rotation.

By contrast, the investigation of precession in time-dependent (as opposed to space-dependent) magnetic fields¹⁴ shows exact reproduction of the effect of rotation; that work left only one question—whether such fields could be found that also satisfy Maxwell's equations—now answered affirmatively. An infinite sheet of uniform current density, starting up from zero to a finite value and returning to zero again, produces a plane-wave electromagnetic 'pulse' whose magnetic field fulfills the requirements. The details of this configuration, with discussions of its possible implementation will be published elsewhere (H.J.B., in preparation).

The deepest issue of interpretation lies in determining when, if ever, an interaction may be considered to produce a symmetry transformation. The issue is joined in noting that torques must be applied to any spinning object to rotate it. These torques imply (certainly for a quantum mechanical system) entanglement with a macroscopic system, such as the magnetic field in the neutron experiments.

Wheeler²³ terms the sign change of a spinor under 2π rotation its 'entanglement relation', referring to the model of a macroscopic object tied (by elastic strings, ribbons or a strip of material) to its surroundings. Under 2π -rotations the strings become irretrievably entangled, but after 4π -rotation they may be topologically restored to the untangled state without rotating the object. This 'entanglement relation' underlies a variety of popularized demonstrations of spinor properties. Dirac was fond of twisting his belt to show it. A triangular lecture-hall model has been described²¹, but perhaps the most beautiful demonstration is that of the Phillipine Binasuan dance²⁴. The causal

connections of interaction may be less surprising than those of correlation (as in recent implementations of the Einstein-Rosen-Podolsky 'paradox'²⁵), but even if they turn out to be pictured or controlled more easily, they are no less real.

I thank Professors J. Byrne and A. Zeilinger for conversations that prompted this paper and A. Zeilinger, M. Horne, C. Shull and D. Gilden for discussion of its first draft. I also thank Professor Shull for providing the hospitality and support of the Massachusetts Institute of Technology Neutron Diffraction Laboratory during the completion of this work, which was partially supported by NSF grant No. 8316209-DMR.

Received 14 October 1984; accepted 28 February 1985.

1. Aharonov, Y. & Susskind, L. *Phys. Rev.* **158**, 1237-1238 (1967).
2. Bernstein, H. J. *Phys. Rev. Lett.* **18**, 1102-1103 (1967).
3. Rauch, H. *et al. Phys. Lett.* **54A**, 425-427 (1975).
4. Rauch, H., Wilfing, A., Bauspiess, W. & Bonse, U. *Z. Phys.* **B29**, 281-284 (1978).
5. Rauch, H. & Zeilinger, A. *Hadr. J.* **4**, 1280-1294 (1981).
6. Werner, S. A., Colella, R., Overhauser, A. W. & Eagen, C. F. *Phys. Rev. Lett.* **35**, 1053-1056 (1975).
7. Klein, A. G. & Opat, G. I. *Phys. Rev. Lett.* **37**, 238-240 (1976).
8. Klemp, E. *Phys. Rev. D13*, 3125-3129 (1976).
9. Stoll, M. E., Vega, A. J. & Vaughan, R. W. *Phys. Rev. A16*, 1521-1524 (1977).
10. Stoll, M. E., Wolff, E. K. & Mehring, M. *Phys. Rev. A17*, 1561-1567 (1978).
11. Kaiser, R. *Can. J. Phys.* **56**, 1321-1332 (1978).
12. Zeilinger, A. *Nature* **294**, 544-546 (1981).

13. Byrne, J. *Nature* **275**, 189-191 (1978).
14. Bernstein, H. J. & Zeilinger, A. *Phys. Lett.* **75A**, 169-172 (1980).
15. Klein, A. G. & Werner, S. A. *Rep. Prog. Phys.* **46**, 259-335 (1983).
16. Price, R. H. *Am. J. Phys.* **50**, 300-329 (1982).
17. Bernstein, H. J. in *Neutron Interferometry* (ed. Bonse, U. & Rauch, H.) 231-240 (Oxford University Press, 1979).
18. Moore, G. T. *Am. J. Phys.* **38**, 1177-1180 (1970).
19. Mezei, F. in *Neutron Interferometry*, 265-272 (Oxford University Press, 1979).
20. Balcar, E. in *Neutron Interferometry*, 252-264 (Oxford University Press, 1979).
21. Bernstein, H. J. *Scientific Research* (NY) **4**, 32-33 (1969).
22. Eder, G. & Zeilinger, A. *Nuovo Cimento B34*, 76-90 (1976).
23. Wheeler, J. A. *Einstein's Vision*, 69-75 and Fig. 70 (Springer, Berlin, 1968).
24. Bernstein, H. J. & Phillips, A. V. *Scient. Am.* **245**(7), 121-137 (1981).
25. Aspect, A. *et al. Phys. Rev. Lett.* **49**, 1804-1807 (1982).

Evidence from polar ice cores for the increase in atmospheric CO₂ in the past two centuries

A. Neftel, E. Moor, H. Oeschger & B. Stauffer

Physics Institute, University of Bern, Sidlerstrasse 5, CH-3012 Bern, Switzerland

Precise and continuous measurements of atmospheric CO₂ concentration were first begun in 1958 and show a clear increase from 315 parts per million by volume (p.p.m.v.)¹ then to 345 p.p.m.v. now. A detailed knowledge of the CO₂ increase since preindustrial time is a prerequisite for understanding several aspects of the role of CO₂, such as the contribution of biomass burning to the CO₂ increase and the sensitivity of climate to the CO₂ concentration in the atmosphere. Estimates of the preindustrial CO₂ concentration are in the range 250-290 p.p.m.v. (ref. 2), but the precise level then and the time dependence of the increase to the present levels remain obscure. The most reliable assessment of the ancient atmospheric CO₂ concentration is derived from measurements of air occluded in ice cores. An ice core from Siple Station (West Antarctica) that allows determination of the enclosed gas concentration with very good time resolution has recently become available. We report here measurements of this core which now allow us to trace the development of the atmospheric CO₂ from a period overlapping the Mauna Loa record back over the past two centuries.

Air bubbles are a characteristic feature of natural ice. At locations with mean air temperatures well below the freezing point, ice is formed by sintering of dry firn. At the firn-ice transition, the pore volume becomes separated as isolated bubbles having no further interaction with the atmosphere. The transition process is slow, so that the sampling of air into isolated bubbles spans at least several years, the duration depending mainly on the accumulation rate and the mean annual temperature.

Measurements were made on a 200-m ice core drilled in the Antarctic summer 1983/84 at Siple Station (75°55' S, 83°55' W) by the Polar Ice Coring Office in Nebraska and our institute. The mean annual air temperature is -24 °C and the annual accumulation rate is 500 kg m⁻². Only one clearly identifiable melt layer of irregular thickness (2-10 mm) was observed in the entire core at 7 m below surface (m.b.s.). Counting the seasonal variations of the electrical conductivity of the ice core allowed us to date the core over the past 200 yr with an accuracy of

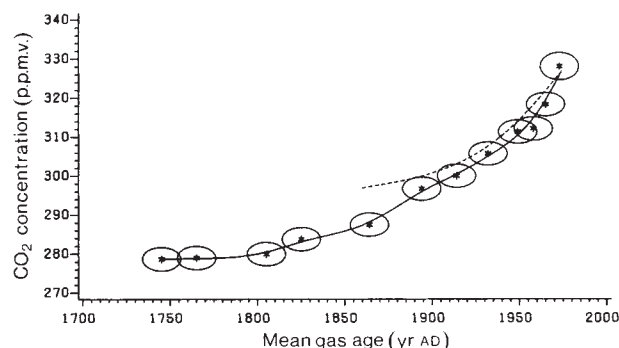


Fig. 1 Measured mean CO₂ concentration plotted against the estimated mean gas age. The horizontal axis of the ellipses indicates the close-off time interval of 22 yr. The uncertainties of the concentration measurements are twice the standard deviation of the mean value, but not lower than the precision of 1% of the measurement. The dotted line represents the model-calculated back extrapolation of the atmospheric CO₂ concentration, assuming only CO₂ input from fossil fuel².

Methods. After crushing the ice samples, the gas expands over a cold trap, condensing the water vapour at -80 °C into the absorption cell. Care was taken to minimize the effect of selective transport of CO₂ by water vapour. This effect was discovered while testing the system, when a difference in the measured CO₂ concentration was observed depending on whether the valve between the crusher and the cold trap was open or closed during the crushing process. With the valve closed, the concentrations were systematically higher by ~15 p.p.m.v. Crushing the ice with the valve open causes small ice particles to be transported to warm surfaces, where they immediately evaporate, leading to a considerable amount of water vapour in the system. The part of this water vapour not immediately condensed in the cold trap expands into the absorption cell and then flows slowly back into the cold trap, as indicated by the pressure monitoring in the absorption cell, which leads to a decreased CO₂ concentration in the absorption cell. When the valve is closed, the ice particles created during the crushing process are kept back in the cold crusher and less water vapour reaches the absorption cell after the valve is opened, leading to smaller deviations of the CO₂ concentration. Also, for this slightly modified measuring procedure, the measurements could be controlled by crushing pure gas-free single-crystal ice surrounded by the same amount of standard gas as contained in an average ice sample. With this procedure, small positive deviations in the CO₂ concentration, attributable to water vapour flow from the crusher into the cold trap, can be determined and the measurements corrected accordingly. This correction is below 1% for 10-cm³ samples and below 4% for 1-cm³ samples. All Siple samples were measured using the modified procedure and each fourth measurement was a calibration run.

Table 1 Mean CO₂ concentration in air extracted from ice samples

Depth below surface (m)	Samples measured	Siple Station		CO ₂ concentration	
		Ice from (yr AD)	Air enclosed (yr AD)	Extracted air (p.p.m.v.)	Atmosphere (p.p.m.v.)
68.2–68.6	8	1891	1962–1983	328 ± 3.5	328
72.4–72.7	11	1883	1954–1976	318 ± 3.0	321
76.2–76.6	11	1876	1947–1969	312 ± 3.0	315
82.0–83.0	28	1867	1938–1960	311 ± 3.0	
92.0–93.0	25	1850	1921–1943	306 ± 3.0	
102.0–103.0	26	1832	1903–1925	300 ± 3.0	
111.0–112.0	26	1812	1883–1905	297 ± 3.0	
128.0–129.0	47	1782	1842–1864	288 ± 3.0	
147.0–147.2	10	1743	1814–1836	284 ± 3.0	
162.0–162.3	9	1723	1794–1819	280 ± 3.0	
177.0–177.3	10	1683	1754–1776	279 ± 3.0	
187.0–187.3	10	1663	1734–1756	279 ± 3.0	
South Pole					
	IRLS/LES			IRLS	LES
139.0–141.0	12/5	820	1660–1880	278 ± 3.0	280 ± 3.0
160.0–162.0	12/5	610	1450–1670	281 ± 3.0	283 ± 3.0
205.0–207.0	8/4	110	950–1170	279 ± 4.0	280 ± 3.0

LES, Large extraction system.

±2 yr (ref. 3). Based on porosity measurements, the time lag between the mean age of the gas and the age of the ice was determined to be 95 yr and the duration of the close-off process to be 22 yr (ref. 4). These values are, of course, evaluated for one particular core representing the present situation (1983), assuming a homogeneous enclosure process and not taking into account the sealing effect of observed impermeable layers.

The gases from ice samples were extracted by a dry-extraction system, in which bubbles are crushed mechanically to release the trapped gases, and then analysed for CO₂ by infrared laser absorption spectroscopy (IRLS) or by gas chromatography (GC). The combination of a needle crusher and IRLS allows us to analyse samples of ice as small as 1 g (ref. 5) (see Fig. 1 legend). During the past 2 yr, a new large dry-extraction device for ice samples up to 1 kg has been built and extensively tested⁶. The ice samples are ground *in vacuo* and the gases collected on a cold finger (20 K) and later analysed by both GC and IRLS. Applying both procedures to neighbouring ice samples gave the same concentration to within 1%. The measurements using the needle crusher, published previously^{5,7,8}, were performed using a slightly modified procedure and exhibited generally lower CO₂ concentrations by 15 p.p.m.v. (see Fig. 1 legend). In 1982 an intercalibration study with the Grenoble laboratory⁷ was performed using the small crusher with the older measuring procedure. Based on our new results, the agreement of the intercalibration must be viewed as a discrepancy, which we will try to resolve in the near future with a new intercalibration series.

Figure 1 shows the mean CO₂ concentration of the 12 depth intervals, measured on the Siple core, plotted against the mean gas age. To extend our data series further back in time, a few samples from the South Pole were also measured (and cross-checked with the large extraction system). These values are listed in Table 1.

Five depth intervals between 82 and 130 m.b.s. of the Siple core were measured with a high spatial resolution. Figure 2a–e shows the individual data series together with the $\delta^{18}\text{O}$ results. A seasonal structure could not be identified, indicating that the CO₂ concentration in the bubbles of the Siple core is homogeneously distributed and is independent of the seasonality of the surrounding ice.

Based on earlier studies of ice samples from locations in Greenland, we expected to measure enhanced CO₂ concentration in the extracted gas, especially in summer layers^{8,9} from locations with a mean annual air temperature above –25 °C. For comparable mean gas ages, the CO₂ concentration measured

on samples from the South Pole (–50 °C) and Siple Station (–24 °C) do not differ within the error limit. Therefore, we conclude that the CO₂ level in the bubbles of the Siple core is not influenced by a temperature effect. This is further supported by the lack of seasonal variations in the five data series measured with high spatial resolution.

Between 68 and 69 m.b.s., the porosity measurements showed two layers with bubble volumes above 80 cm³ kg^{–1} ice, which correspond to complete enclosure. This is clearly above the 40 cm³ kg^{–1} ice expected from a smooth homogeneous enclosure process. These impermeable layers have a sealing effect if they are imbedded in firn with reduced permeability¹⁰. Because of the layers between 68 and 69 m.b.s., the air below is already completely isolated, about 7 m above the depth obtained, assuming a homogeneous enclosure. Consequently, for this core, the difference between ice and mean gas age is only 80–85 yr instead of 95 yr as estimated previously⁴.

The enclosure of air occurs in winter layers at shallower depth than in summer layers, which could lead to a seasonal modulation of the CO₂ concentration in the bubbles during a period of changing atmospheric CO₂ concentration in case of a smooth and homogeneous enclosure process. The abovementioned lack of a seasonal variation in the series from 82 m.b.s. therefore further supports the sealing effect of such layers.

If the 328 p.p.m. measured at a depth of 68.5 m.b.s. is matched with the atmospheric South Pole record¹¹, the mean gas age is 10 yr, corresponding to a difference between mean gas age and ice age of 82 yr, which lies in the above estimated range. This difference is used in calculating the mean gas age for all depths. That the CO₂ concentration measured on the subsequent samples from 72.5 and 76.5 m.b.s. corresponds with the atmospheric South Pole record justifies this age determination and indicates, also, that impermeable layers at a depth of ~68 m.b.s. occur frequently.

There remains the question of the duration of the gas-enclosure time in such an impermeable layer. Assuming, as a first approximation, a constant and continuous enclosure rate ending at the time of core drilling ($t_2 = \text{AD}1983$), the beginning (t_1) can be calculated from the expression $(t_2 - t_1)C_{\text{bubbles}} = \int_{t_1}^{t_2} C_{\text{atm}}(t) dt$, where $C_{\text{atm}}(t) = a + b(t - 1958) + c(t - 1958)^2$ (with t in yr AD). The atmospheric concentration, C_{atm} , is taken from a quadratic fit of the measured CO₂ data in the atmosphere. The time t_1 so determined is 1963, so that the duration of the close-off is 20 yr, in good agreement with the estimate of Schwander and Stauffer⁴.

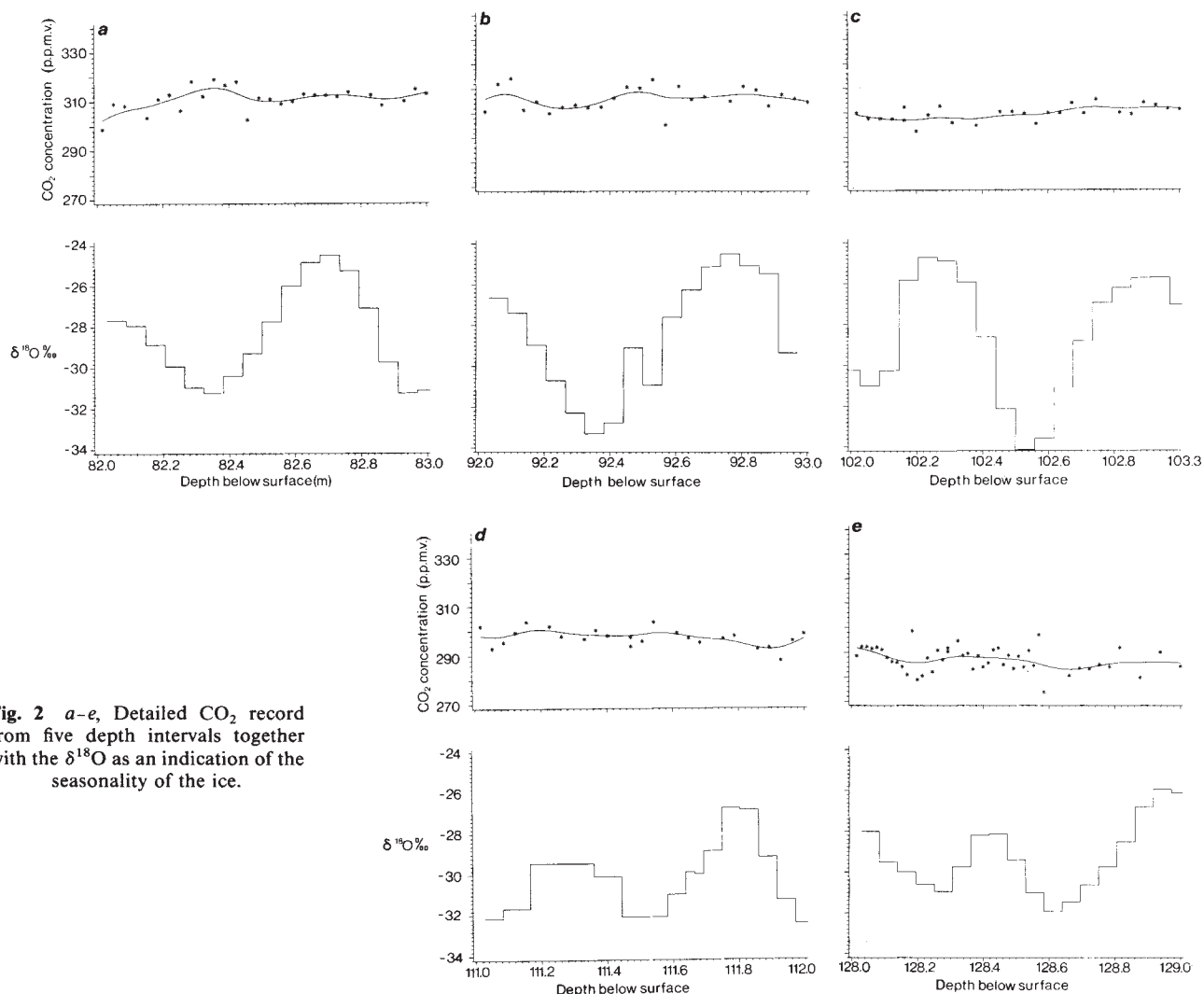


Fig. 2 a-e, Detailed CO₂ record from five depth intervals together with the δ¹⁸O as an indication of the seasonality of the ice.

We would have more difficulties explaining values lower by 15 p.p.m.v., as obtained, for example by the open-valve method. This would require a 20–40-yr-old mean gas age in the bubbles, depending on the development of the atmospheric CO₂ concentration. A very much prolonged close-off interval, a possibility which we exclude⁴, or an incomplete mixing of air in the firn would be required. Also, the atmospheric CO₂ concentration would have to be increased between 1930 and 1960 by ~ 1 p.p.m.v. yr⁻¹, a similar rate to that of the past 20 yr.

Based on these results, we conclude that the atmospheric CO₂ concentration around 1750 was 280 ± 5 p.p.m.v. and has increased since, essentially because of human factors, by 22.5% to 345 p.p.m.v. in 1984. Back extrapolation of the Mauna Loa curve, on the basis of the fossil fuel input data obtained by Rotty¹² and assuming zero CO₂ input from biomass burning, suggests a preindustrial level of 297 p.p.m.v. (ref. 2). The difference of 17 p.p.m.v. between this calculated value and that

measured indicates a significant contribution from biomass burning, even during the past century when the contribution from fossil fuel consumption was still small. The CO₂ increase in the Siple core will be studied in more detail using δ¹³C measurements. Preliminary δ¹³C results on an ice core from the South Pole are indeed very promising¹³. The CO₂/air ¹³C/¹²C record on the Siple ice core and the ¹⁴C/¹²C Suess effect from tree rings, together with the fossil fuel input-time history, will then enable us to separate the CO₂ inputs from those attributable to fossil fuels and biomass. It will also allow us to determine the effective airborne fraction for the combined input and to check the validity of the present carbon-cycle models.

We thank Drs J. Schwander, W. Bernhard, K. Hänni and R. Walther for assistance with the field work and with the analysis in the laboratory. We also thank D. Raynaud for reviewing the manuscript. This work was supported by the Swiss National Foundation, the Department of Polar Programs of the NSF, the US Department of Energy and the University of Bern.

Received 31 December 1984; accepted 15 March 1985.

1. Keeling, C. D. *et al. Tellus* **28**, 538–551 (1976).
2. WMO Rep. 10 (World Meteorological Organization, Geneva, 1983).
3. Schwander, J. thesis, Univ. Bern (1984).
4. Schwander, J. & Stauffer, B. *Nature* **311**, 45–47 (1984).
5. Zumbun, R., Neftel, A. & Oeschger, H. *Earth planet. Sci. Lett.* **60**, 318–324 (1982).
6. Moor, E. & Stauffer, B. *J. Glaciol.* (in the press).
7. Barnola, J. M., Raynaud, D., Neftel, A. & Oeschger, H. *Nature* **303**, 410–412 (1983).

8. Neftel, A., Oeschger, H., Schwander, J., Stauffer, B. & Zumbun, R. *Nature* **295**, 220–223 (1982).
9. Stauffer, B. *et al. Ann. Glaciol.* **5**, 160–164 (1984).
10. Stauffer, B. *et al. Ann. Glaciol.* (in the press).
11. Keeling, C. D. *et al. Tellus* **28**, 552–564 (1976).
12. Rotty, R. M. in *Carbon Cycle Modelling* Scope 16 (ed. Bolin, B.) 121–126 (Wiley, New York, 1981).
13. Friedli, H. *et al. Geophys. Res. Lett.* **11**, 1145–1148 (1984).

Vertical structure of the Brazil Current

David L. Evans

Graduate School of Oceanography, University of Rhode Island,
Kingston, Rhode Island 02881, USA

Sergio S. Signorini*

Universidade de Sao Paulo, Instituto Oceanografico, Sao Paulo,
CEP-05508, Brazil

Recent interest in world climate and interaction of the ocean and atmosphere have led to studies of the meridional fluxes of fresh water and heat in the ocean^{1,2}. Several such studies^{3,4} have noted the asymmetry of heat fluxes between the North and South Atlantic oceans. In particular, the fluxes across both 24° N and 24° S appear to be northward. Crucial to all of the direct calculation techniques is an accurate estimate of the transport of the western boundary currents. In the North Atlantic, useful measurements in the Florida Current⁵ over a long period of time are available. For the much less studied South Atlantic no comparable time series has been made. The direct measurements of Brazil Current velocities near 23° S reported here show southward flow of warm water above 400 m depth and northward flow of Antarctic Intermediate Water below that. The upper layer transport was $\sim 6 \times 10^6 \text{ m}^3 \text{ s}^{-1}$ towards the south offshore of the 200-m isobath, with an indication of comparable flow on the shelf.

The Brazil Current is the South Atlantic analogue of the Gulf Stream. For the present purposes, it is defined as the southward-directed western boundary current, extending downward from the surface, closing the circulation of the subtropical, wind-driven circulation in the South Atlantic. From a consideration of the wind-driven anticyclonic gyres, the two should be quite similar. As Stommel⁶ points out, "therefore ... the wind-driven component of the circulation is much the same in the two hemispheres, but ... in the western boundary currents the thermohaline circulation plays quite different roles in the Gulf Stream and Brazil Current, making the former appear much more strongly on the charts of the topography of surfaces, equal pressure, or density". Indeed, the limited historical hydrographical observations indicate that the total transport of the Brazil Current may be no larger than $9.4 \times 10^6 \text{ m}^3 \text{ s}^{-1}$ in the upper 600 m (refs 7, 8). However, as Stommel remarks, the wind-driven component should be as large as $30\text{--}40 \times 10^6 \text{ m}^3 \text{ s}^{-1}$ to balance the equatorward transport in the gyre. Recent calculations based on balancing the meridional transport due to the wind stress curl^{9,10} indicate that up to $30 \times 10^6 \text{ m}^3 \text{ s}^{-1}$ should flow in the Brazil Current near 30° S and $\sim 20 \times 10^6 \text{ m}^3 \text{ s}^{-1}$ should flow at 25° S.

* Present address: E.G.&G. Corporation, Waltham, Massachusetts, USA.

There are longstanding beliefs that western boundary currents provide an essential means for meridional transport and that the Brazil Current region may be the major conduit for the exchange of at least four major water masses between the hemispheres¹¹. In contrast, inverse calculations indicate that the Brazil Current may be important only south of 25° S. In fact, these calculations indicate little southern transport along the boundary north of 30° S (ref. 4). Previous work has relied on water mass boundaries or more arbitrary fixed depths to determine levels of no motion for calculation of geostrophic transport. There have been no systematic investigations of the Brazil Current including direct velocity or transport measurements.

The choice of reference level for geostrophic calculations leads to an ambiguity in even the direction of flow of some of the water masses along the boundary. Traditionally, water mass analysis suggested a layered current system with warm tropical water flowing southward over northward Antarctic Intermediate Water which is in turn above the southward-flowing North Atlantic Deep Water¹². More recent interpretation of the data has led some investigators to argue that the wind-driven gyre penetrates deep enough to carry Intermediate Water to the south after circulating around the basin^{4,13}. Resolution of the issues of transport of the Brazil Current and its vertical structure, that is, direction of flow of the various water masses, motivated the present study. The results presented below indicate that the 'Brazil Current' should be reserved for the flow shallower than $\sim 400 \text{ m}$ at this latitude.

A cruise aboard the *Prof. Besnard* of the Institute of Oceanography at the University of Sao Paulo during April 1982 allowed a detailed expendable bathythermograph (XBT) survey and hydrographic study between 19° S and 24° S to be performed. Essentially, the Brazil Current was found to be closely confined to the region near the continental slope. It flowed continuously past the seamount chain at 20°30' S through the most inshore passage and onto the continental shelf. Geostrophic transports were $< 10 \times 10^6 \text{ m}^3 \text{ s}^{-1}$, using either 500 m or 1,000 m as a reference level¹⁴; they were $3.8 \times 10^6 \text{ m}^3 \text{ s}^{-1}$ and $6.8 \times 10^6 \text{ m}^3 \text{ s}^{-1}$ respectively.

The first direct velocity measurements of the Brazil Current were made in April 1983 aboard RV *Oceanus*. The XBT survey from 19° S to 24° S was repeated in greater detail and extended further offshore than in April 1982. The Pegasus velocity profiler¹⁵ was deployed at two passages in the seamount chain at 20°30' S and along five sites across the Brazil Current near Cabo Frio between 23° S and 24° S to obtain vertical profiles of horizontal velocity with an estimated accuracy of $\pm 2 \text{ cm s}^{-1}$. Figure 1 shows the survey area and location of XBT lines and Pegasus sites. Also shown is the depth of the 15°C isotherm, the gradient of which is indicative of the strength of the baroclinic component of the flow.

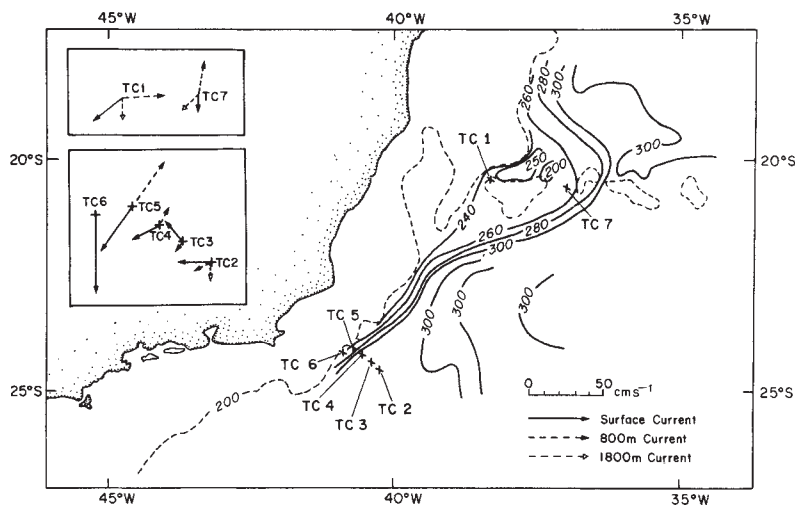


Fig. 1 Survey area for April 1983 cruise showing depth of the 15°C isotherm and locations of Pegasus stations. Also shown are mean velocity vectors at depths characteristic of surface, intermediate and deep waters.

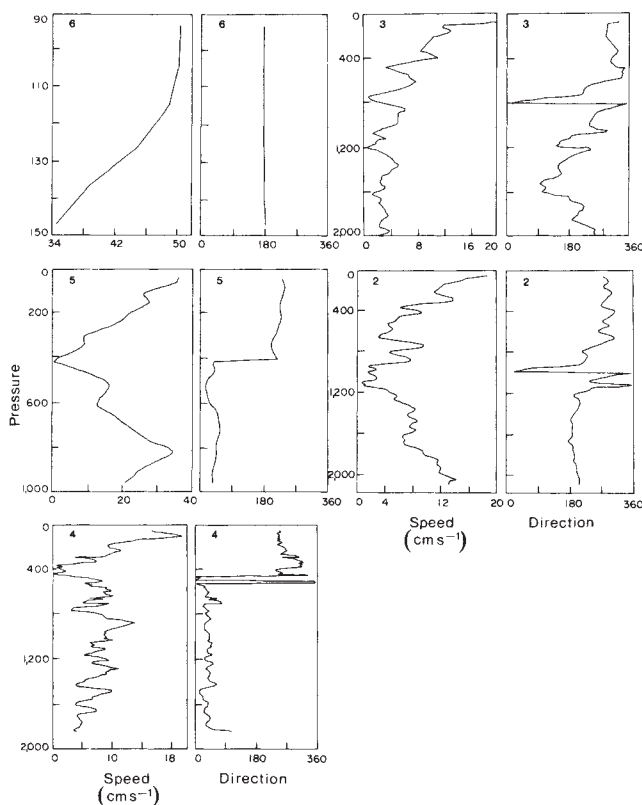


Fig. 2 Vector-averaged profiles of speed and direction at each Pegasus site. Directions are measured as degrees true towards which the current flows. (Note the difference in pressure and speed scales among the frames. Depths are in decibars.)

Offshore from the 2,000-m isobath and near the Abrolhos Banks, the temperature gradient field is poorly organized and there is no clear indication of baroclinic flow. Based on the temperature gradients alone, the Brazil Current seems to be flowing through the banks with weaker intensity than along the 200-m isobath near Cabo Frio, where it is confined to flow in a narrow stream along the shelf-break. This pattern is essentially the same as found in April 1982 except that there is a suggestion of flow to the east of the banks at 20°30' S. Earlier work^{7,8,14} has shown that a close relationship between temperature (T) and salinity (S) in this region permits inference of geostrophic flow from the temperature field to within an accuracy of $\pm 10\%$.

Each of the Pegasus sites TC2–TC6 was visited three times during a period of 10 days. Only two measurements were made at TC1 and one at TC7. Although the profiles over the course of the cruise showed small variability, they represent only one snapshot of a field which is likely to have large variability at longer timescales. We may hope that, like the Gulf Stream, the vertical structure is relatively stable. Figure 2 displays the vector-averaged profiles of speed and direction for the section near Cabo Frio¹⁵. The east and north components of velocity were interpolated to common depths at approximately the original resolution using cubic splines. Vector averages of the three drops at each site were then computed. Figure 3 exhibits all three profiles from site TC5 to illustrate the relatively small variability among the drops.

Site TC2 was selected, based on the isotherm pattern from the XBT survey, to be offshore of the Brazil Current. The surface flow is $< 20 \text{ cm s}^{-1}$ and is directed to the west (or north-west) down to 1,000 m. Between 1,000 and 1,200 m the velocity vector is small ($< 3 \text{ cm s}^{-1}$) and rotates through 360° with depth. Such flow may be characteristic of the subtropical gyre circulation to the east of the boundary current. However, an XBT line along 36° W failed to indicate significant zonal flow and the surface flow does not indicate an eastern source (Fig. 1). From 1,400 m to the bottom, the flow is directed almost due south. Deep southward flow in the absence of a similar surface current is

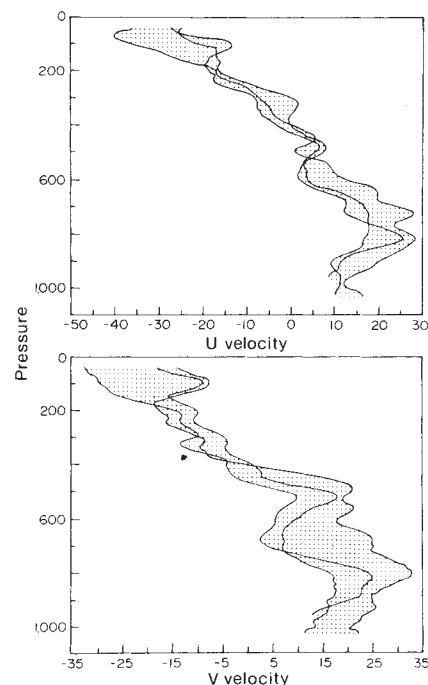


Fig. 3 East and north components of three drops at site TC5. Data were collected over 1 week.

not taken to be indicative of the Brazil Current, but is likely to be part of the deep thermohaline mode. At TC3, the next station inshore, all speeds are $< 20 \text{ cm s}^{-1}$. Surface layer flow is directed to the north-west down to a depth of 500 m. Below 600 m, speeds are quite weak and directions are consequently not well determined. TC4 is the first station definitely in the Brazil Current. In the layer from the surface to 400 m, the current is directed towards the south-west at speeds up to 20 cm s^{-1} . Below 500 m, the water moves northeastward at speeds near 10 cm s^{-1} . TC5 again shows the dramatic change in direction near 400 m depth. The upper layer (traditional Brazil Current water) moves to the south-west at speeds up to 35 cm s^{-1} . Below 400 m, the flow is directed towards the north-east, also at speeds of up to 35 cm s^{-1} . (The salinity minimum, indicative of Intermediate Water, is found near 800 m at this location.) TC6 is from the station made on the shelf in 200 m water depth. This station represents the 'pure' Brazil Current. All flow is directed towards the south-west with speeds decreasing from 50 cm s^{-1} at the surface to 35 cm s^{-1} near the bottom. Ship drift records and navigational charts (from the US Navy Hydrographic Office) indicate that this flow may extend over the entire shelf at comparable speeds.

Data from sites TC1 and TC7 near the seamounts are more difficult to interpret because of the possible influence of the bottom topography. However, both sites show southward surface flow, northward flow of Intermediate Water and southward flow of Deep Water. Note that these data are based on a single profile, not the average of three like those at TC2–TC6. TC1 was revisited at the end of the cruise and speeds were smaller and somewhat rotated from those in Fig. 1. (Surface flow was more westerly and 800 m flow more northerly.) The general sense of the flow of the water masses was, however, the same. All of these profiles support Wüst's mass analysis indicating flow away from known sources, in contrast to the more recent^{11,12} view for the Intermediate Water. Note, however, that these measurements do not preclude the possibility that Intermediate Water is transported towards the west, south of 24° S and then turns equatorward. The significance of the direct measurements for dynamic computations is striking. For example, a choice of 900 m as a reference level at TC5 would yield southward surface speeds of 70 cm s^{-1} (twice the observed) and the wrong direction for the Intermediate Water. Thus charts of dynamic topography relative to 1,000 m may be inappropriate for the study of the Brazil Current at 24° S and attention should be paid to the distinct water mass characteristics.

Using these observations to estimate the transport of the Brazil Current down to 400 m yields $\sim 6 \times 10^6 \text{ m}^3 \text{ s}^{-1}$ towards the south offshore of the 200-m isobath. If TC6 is representative of the flow on the shelf, another $5 \times 10^6 \text{ m}^3 \text{ s}^{-1}$ may flow inshore of the 200-m isobath. Indeed, using the historical T - S relation for the area¹⁴, geostrophic transport computed from temperatures measured by Pegasus is $\sim 5.5 \times 10^6 \text{ m}^3 \text{ s}^{-1}$ relative to 400 m. While the total transport ($11 \times 10^6 \text{ m}^3 \text{ s}^{-1}$) may not then be significantly different from earlier estimates, its vertical distribution is quite dissimilar in being confined to the upper 400 m. Estimates of the poleward heat flux will also differ due to the higher mean temperature of the water moving towards the pole—an effect which is enhanced by the strong equatorward flow of cooler Antarctic Intermediate Water.

This work was supported in part by NSF grant OCE82-14834. L. Miranda and A. Mascarenhas provided valuable assistance in collecting the data. Logistical support from the University of Sao Paulo was essential to the field programme.

Received 27 November 1984; accepted 11 March 1985.

1. Bryden, H. & Hall, M. *Science* **207**, 884 (1980).
2. Hastenrath, S. *J. phys. Oceanogr.* **10**, 159 (1980).
3. Stommel, H. *Proc. natn. Acad. Sci. U.S.A.* **77**, 2377 (1980).
4. Fu, L. *J. phys. Oceanogr.* **11**, 1171 (1981).
5. Niiler, P. & Richardson, W. *J. mar. Res.* **31**, 144 (1973).
6. Stommel, H. *The Gulf Stream* (University of California Press, 1958).
7. Signorini, S. *Deep-Sea Res.* **25**, 481 (1978).
8. Miranda, L. & Filho, B. *Cienc. interam.*, 44-48 (1981).
9. Evenson, A. & Veronis, G. *J. mar. Res.* **33**, 131 (1975).
10. Hellerman, S. & Rosenstein, M. *J. phys. Oceanogr.* **13**, 1093 (1983).
11. Reid, J., Nowlin, W. & Patzert, S. *J. phys. Oceanogr.* **7**, 62 (1977).
12. Wüst, G. in *Wissenschaftliche Ergebnisse der Deutschen Atlantischen Expedition auf dem Forschungs- und Vermessungsschiff "Meteor"* 6, Pt 1, 2 (1935).
13. Buscaglia, J. *J. mar. Res.* **29**, 245 (1971).
14. Evans, D., Signorini, S. & Miranda, L. *J. phys. Oceanogr.* **13**, 1171 (1983).
15. Spain, P., Dorson, D. & Rossby, T. *Deep-Sea Res.* **28A**, 1553 (1981).

The Somali Current in autumn 1984, before the onset of the north-east monsoon

Friedrich Schott* & Michele Fieux†

* Rosenstiel School of Marine and Atmospheric Science, University of Miami, Miami, Florida 33149, USA

† Laboratoire d'Océanographie Physique du Muséum National d'Histoire Naturelle, 75231 Paris, France

The Somali Current undergoes vigorous variations during the course of the south-west monsoon. Although the initial phases of this response were fairly intensively studied during the FGGE Indian Ocean Experiment (INDEX) in 1979^{1,2}, relatively little is known about what happens in the late summer monsoon and the transition to the winter monsoon. In October 1984, we carried out measurements in the Somali Current regime with a shipboard acoustic Doppler current profiler which showed that a continuous Somali Current did not exist at this time; instead, the cross-equatorial flow with a transport of $12 \times 10^6 \text{ m}^3 \text{ s}^{-1}$ in the upper 100 m identified by relatively low surface salinities, was found to turn offshore south of 2° N . Between 6 and 11° N , a large closed anticyclonic gyre, marked by relatively high surface salinities, was found with a transport of $32 \times 10^6 \text{ m}^3 \text{ s}^{-1}$ in the upper 250 m, which is significantly greater than previously estimated from geostrophic calculations.

In the initial response to the onset of the south-west monsoon, from early May to early June, the Somali Current develops as a shallow cross-equatorial current, turning offshore at about 3° N with a cold upwelling wedge at the turnoff latitude; after the onset of the strong wind stress curl off Somalia in mid-June, a second gyre is generated between 5 and 10° N with a second cold wedge at the coast on its northern end. Satellite sea-surface temperature and ship measurements indicate that at some time between late July and September the southern cold wedge starts to propagate northward and the two-gyre pattern seems to

collapse^{2,3}. Ship measurements in different years showed that the water behind the northward-travelling southern cold wedge was low-saline water originating south of the Equator, replacing the higher-saline water found between 5 and 10° N in May to July^{2,4,5}. The result of this development is that in the late summer monsoon, the Somali Current is a continuous boundary current from south of the Equator to the Horn of Africa. It is not known in any detail what happens in the transition period between the decay of the south-west monsoon in late August and the onset of the winter monsoon in early November.

Nothing is known about the cross-equatorial flow at this time, partly because geostrophy cannot be applied close to the Equator to calculate transports from hydrographical data. For the northern Somali Current region, expendable bathythermograph (XBT) measurements (for example, refs 6, 7) suggest that a northern gyre continues in that period, but its water origin, current structure, transports and any relation of this gyre to the current further south are not really known.

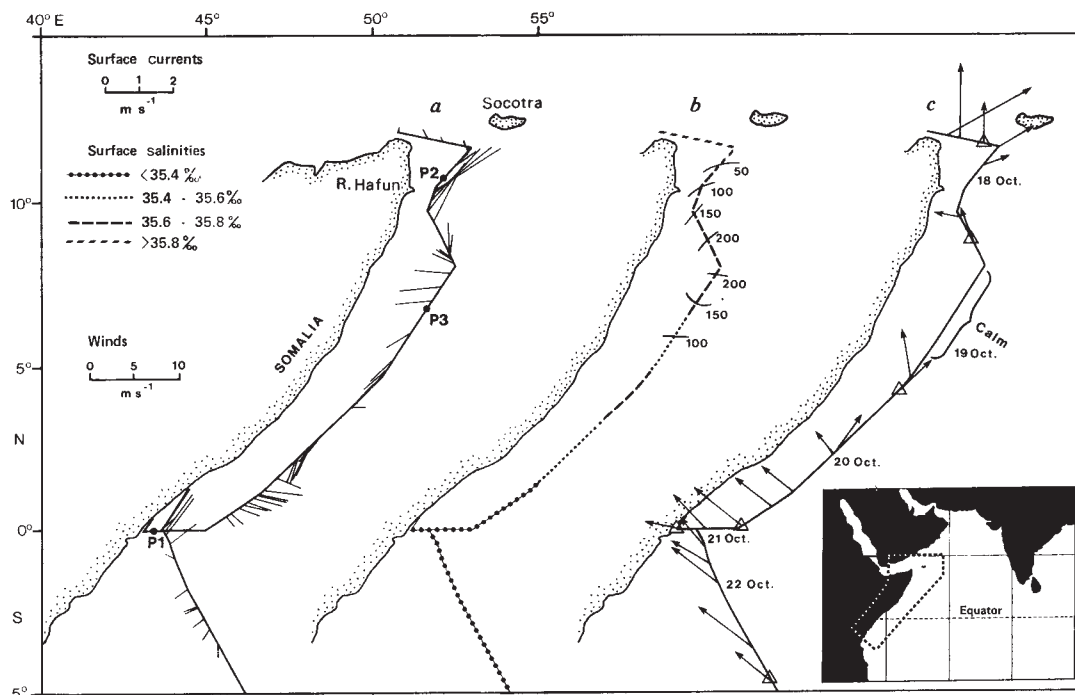
During part of this transition period, from 17 to 22 October 1984, we performed measurements in the Somali Current system on board the French RV *Marion Dufresne* with a shipboard acoustic Doppler current profiler, XBTs and continuous surface-temperature and salinity recordings (Fig. 1).

The cross-equatorial flow in the Somali Current in October 1984, as the surface-current vectors (Fig. 1a) and the surface salinities of 35.0–35.2‰ (Fig. 1b) indicate, does not continue northward but turns offshore south of about 2° N . Its origin is the South Equatorial Current, the biggest branch of which forms a northward-flowing boundary current east of Madagascar, which then continues westward around the northern tip of Madagascar across to Africa, where it is called the East African Coast Current. Through admixtures of the low-saline coastal waters from East African rivers, the surface salinity is slightly reduced relative to that of the inflowing South Equatorial Current.

The northward current component through the equatorial section (Fig. 2b) has a subsurface maximum of more than 150 cm s^{-1} near the coast at about 40 m depth and still velocities of more than 50 cm s^{-1} at 200 m depth. The northward transport through that section in the upper 100 m is $12.3 \times 10^6 \text{ m}^3 \text{ s}^{-1}$, which is almost exactly matched by the transport turning offshore in the top 100 m between 2° N and the Equator, for which we found $12.6 \times 10^6 \text{ m}^3 \text{ s}^{-1}$ from the acoustic velocity-profile measurements. The cross-equatorial transport in the upper 100 m is comparable to the transports determined by Leetmaa *et al.*¹ during five repetitions of a transport section normal to the coast at 1° S in the period May–June 1979, for which they obtained a mean transport of $14.0 \times 10^6 \text{ m}^3 \text{ s}^{-1}$. Even the northward flow in the depth range 100–300 m for which we obtained $8.4 \times 10^6 \text{ m}^3 \text{ s}^{-1}$ (the quality of the acoustic current-profile data deteriorates below $\sim 250 \text{ m}$) is comparable in magnitude to that of the 1979 early-summer measurements which showed a mean of $6.0 \times 10^6 \text{ m}^3 \text{ s}^{-1}$ (ref. 1). In summary, the cross-equatorial transport of the Somali Current in late autumn is very similar to that during the onset of the south-west monsoon. The local winds off Kenya/Somalia are quite different during May/June and October (Fig. 1c), whereas the trade winds over the subtropical south Indian Ocean are quite similar in these two phases, which suggests that the cross-equatorial flow in the early phase of the south-west monsoon is driven primarily by remote forcing through the inflow from the South Equatorial Current, not by the local monsoon winds off East Africa.

North of 6° N , the ship-drift vectors (Fig. 1a) indicate a large anticyclonic gyre, which had surface salinities of 35.6–35.8 parts per thousand (p.p.t.) (Fig. 1b), that is, significantly higher than in the cross-equatorial current. Similar salinity differences between cross-equatorial flow and the northern regime along that longitude were found in earlier autumn cruises with the *Marion Dufresne*⁸. The current in the northern branch of this gyre was surface intensified (Fig. 2a) despite the lack of strong local winds during that time (Fig. 1c). Transport of this gyre in the upper 250 m was $33.5 \times 10^6 \text{ m}^3 \text{ s}^{-1}$ westward between 6° and

Fig. 1 Measurements along the track of RV *Marion Dufresne* during 18–22 October 1984. The Doppler profiler is an Ametek-Straze Model DCP 4400 operating at 115 kHz. Good data were obtained down to about 250 m depth or more at quiet to moderate sea state. Data storage at sea was in averages of 10 profiles at a time interval of 32 s for the averages. Subsequently, for the purpose of transport calculation, data were averaged to 15-min profiles, that is, over 280 individual profiles. In earlier comparisons in the Florida Current with other direct current measurements, we had determined that for sample sizes of this magnitude the accuracy of the acoustically measured horizontal currents, relative to the ship, was a few centimetres per second. The ship's motion was determined using satellite positioning. *a*, Drift vectors (P1–P3, locations of profiles shown in Fig. 2a) determined from the ship's satellite navigation system and the two-component electromagnetic speed log. However, satellite passes occur irregularly and gaps of several hours between fixes actually occurred in that region. Although this is unsatisfactory for calibration of individual profiles, it may be sufficient for the total transport through some section, say the equatorial section of the ship track, because for that calculation only the positions at both ends of the section need to be known. This means that the longer the section, the less the importance of position inaccuracies. *b*, Surface salinities and depth of 20 °C isotherm north of 5° N. *c*, Wind vectors, also indicating midnight ship positions (Δ , 00Z positions). Inset (lower right) shows area of observations off north-east Africa.



a, Drift vectors (P1–P3, locations of profiles shown in Fig. 2a) determined from the ship's satellite navigation system and the two-component electromagnetic speed log. However, satellite passes occur irregularly and gaps of several hours between fixes actually occurred in that region. Although this is unsatisfactory for calibration of individual profiles, it may be sufficient for the total transport through some section, say the equatorial section of the ship track, because for that calculation only the positions at both ends of the section need to be known. This means that the longer the section, the less the importance of position inaccuracies. *b*, Surface salinities and depth of 20 °C isotherm north of 5° N. *c*, Wind vectors, also indicating midnight ship positions (Δ , 00Z positions). Inset (lower right) shows area of observations off north-east Africa.

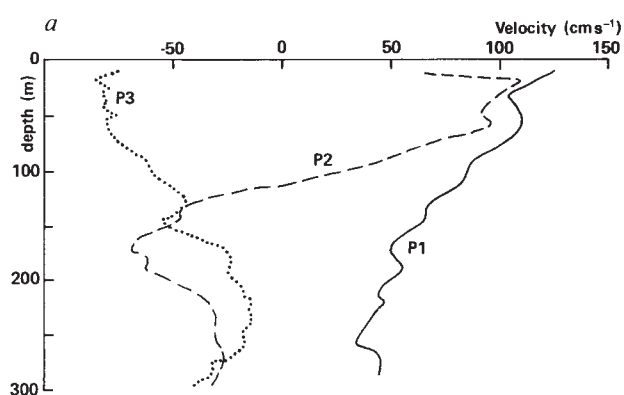
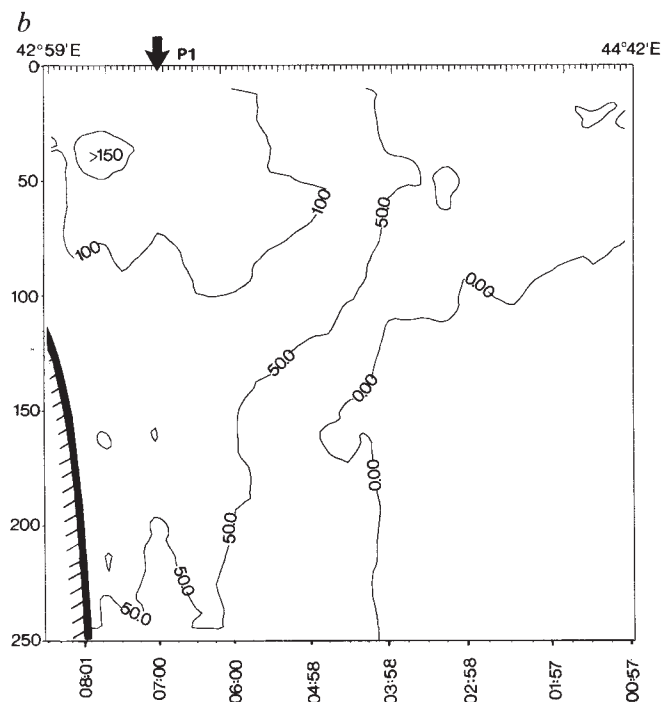


Fig. 2 *a*, Profiles of north component of current at point P1 (0°02'N, 43°19'E) in cross-equatorial flow and of east components at points P2 (10°40'N, 52°04'E) and P3 (6°43'N, 51°30'E) in the northern gyre (Fig. 1a). *b*, Section of the northward current component along the Equator (for location, see Fig. 1a) for 21 October 1984 (start 00:57:05; stop 08:45:40).



8.5° N as determined from the acoustic Doppler current profiles, and $31.6 \times 10^6 \text{ m}^3 \text{ s}^{-1}$ eastward between 8.5° and 11.5° N. This is a closed mass budget within a few per cent. This is surprising because it has to be kept in mind that these transport numbers can be significantly affected by a small misalignment of the horizontal axis of the acoustic transducer. We corrected for the misalignment by comparing the vectors from satellite and radar navigation with those from acoustic bottom tracking near the coast, but an error of several $10^6 \text{ m}^3 \text{ s}^{-1}$ is still possible when integrating over long track segments as done here. It appears though, from the good balance both in the northern gyre and the cross-equatorial flow, that the compensation was effective. Earlier geostrophic transport calculations⁷ obtained $22 \times 10^6 \text{ m}^3 \text{ s}^{-1}$ in the 0–400-m layer for an XBT section in October, from which densities and geostrophic currents were derived using a mean temperature/salinity relation. When applying the

same method⁷ to our XBT data, we obtain $17 \times 10^6 \text{ m}^3 \text{ s}^{-1}$, both significantly lower than our direct measurements for just the upper 250 m. One reason for the difference is that the current at the assumed reference level of 400 m is not zero; another reason is that there is significant deviation from plain geostrophy through the curvature term.

Current profiles in the eastward- and westward-flowing branches of the northern gyre (Fig. 2a) show that the northern gyre is not symmetrical, which is also obvious from the ship drifts (Fig. 1a) and the depths of the 20 °C isotherms (Fig. 1b).

This asymmetry, which is typical and also is visible in earlier XBT sections through the northern gyre^{6,7}, can be explained by the strong nonlinearity of that circulation system, which causes a shift of high currents into the northwestern corner (for example, ref. 9.)

As mentioned earlier, the Somali Current seems usually to go through a phase (in August–September after the collapse of the two-gyre system where the Somali Current is a continuous boundary current) of carrying low-saline water from south of the Equator to $\sim 10^\circ \text{N}$ ^{2,4,5}. If this situation also occurred in 1984, we conclude from our observations that the current of low-saline water must retract again in late September–mid-October. Where, then, is all the low-saline water which was found all the way up to Ras Hafun in $\sim 200\text{-m}$ -thick layer between the coast and $\sim 300\text{ km}$ offshore in late August in these earlier studies? The transport of this water was estimated at $13 \times 10^6 \text{ m}^3 \text{ s}^{-1}$ by Schott¹⁰, about equivalent to the cross-equatorial transport found in the early south-west monsoon phase by Leetmaa *et al.*¹. Another possibility is that in 1984 the two-gyre pattern did not collapse during August/September but remained intact from June to October. It is not known whether

this can actually happen, but the possibility will be investigated using satellite surface-temperature data.

It is interesting that only very small currents were observed between Socotra and Cap Guardafui (Fig. 1a). This is at variance with recent numerical model results of Luther and O'Brien¹¹ and G. Philander (personal communication), who, although using quite different models, both show the northern gyre to propagate further northward and to circulate around Socotra in mid-October, that is, with its full transport through the passage between Socotra and the mainland.

A technical conclusion from this first application of the Ametek Doppler profiler in the Indian Ocean is that the system has provided very useful data of the near-surface circulation, which could not be easily obtained by any other means.

We thank the Terres Australes et Antarctiques Françaises (TAAF) for having made this cruise possible and the captain and crew of RV *Marion Dufresne* for their enthusiasm and help with the shipboard work. We also thank P. Bedard, A. Kartavtseff, B. Olivier and G. Samuels for help in acquisition and processing of the data. This study was supported by the Office of Naval Research under contract N-00014-80C-0042.

Received 21 January; accepted 19 March 1985.

1. Leetmaa, A. Quadfasel, D. R. & Wilson, D. J. *phys. Oceanogr.* **12**, 1325–1342 (1982).
2. Swallow, J. C., Molinari, R., Bruce, G., Brown, O. B. & Evans, R. H. *J. phys. Oceanogr.* **13**, 1398–1415 (1983).
3. Evans, R. H. & Brown, O. B. *Deep-Sea Res.* **28**, 521–527 (1981).
4. Warren, B. A., Stommel, H. & Swallow, J. C. *Deep-Sea Res.* **13**, 825–860 (1966).
5. Bruce, J. G. *Deep-Sea Res.* **20**, 837–846 (1973).

6. Bruce, J. G. *J. geophys. Res.* **84**, 7742–7748 (1979).
7. Bruce, J. G., Feix, M. & Gonella, J. *Oceanol. Acta* **4**, 7–9 (1981).
8. Feix, M. & Levy, C. in *Hydrodynamics of the Equatorial Ocean* (ed. Nihoul, J. C.) 17–29 (Elsevier, Amsterdam, 1983).
9. Bryan, K. J. *atmos. Sci.* **20**, 594–606 (1963).
10. Schott, F. *Prog. Oceanogr.* **12**, 357–381 (1983).
11. Luther, M. & O'Brien, J. *Prog. Oceanogr.* **14**, 353–385 (1985).

Determination of the age of the Australian continent by single-grain zircon analysis of Mt Narryer metaquartzite

Urs Schärer* & Claude J. Allegre

Laboratoire de Géochimie et Cosmochimie, Institut de Physique du Globe et Département des Sciences de la Terre, Universités Paris VI et VII, 4 Place Jussieu, 75230 Paris Cedex 05, France

A fundamental constraint on the evolution of the Earth is the time of the first occurrence of continental crust. The recent publication¹ of ages of 4,200–4,100 Myr for zircons from Australia is therefore of considerable importance. This dating was achieved by single-grain analysis on the SHRIMP ion microprobe at the Australian National University. To substantiate the existence of these old zircons by an alternative method, we studied the same zircon population by conventional solid-source mass spectrometry and isotope dilution analysis of single grains and individual fragments of single grains. Our results confirm neither the 4,200–4,100-Myr ages nor the particular discordancy patterns obtained on the ion probe. The age pattern we obtained can be explained by zircon crystallization 3,800–3,300 Myr ago. Probable primary ages of $\sim 3,750$ – $3,600$ Myr coincide with model Sm/Nd ages for the adjacent gneisses², the possible source rocks of the Mt Narryer quartzite, and the younger ages are interpreted as those of zircon overgrowth compatible with a Rb/Sr age for the regional metamorphic overprint².

In their ion-probe study, Froude *et al.*¹ detected, among the 20 precisely analysed grains, 4 grains with apparent minimum ages older than 4,000 Myr, whereas the others yielded minimum ages between about 3,100 and 3,750 Myr. However, the data for the latter have not yet been published, which precludes a direct comparison with the results presented here. We have analysed 27 single zircon grains and 12 individual fragments of another 5 single grains. Comments on the isotope dilution technique are given in the footnotes to Table 1, which lists the U/Pb analytical

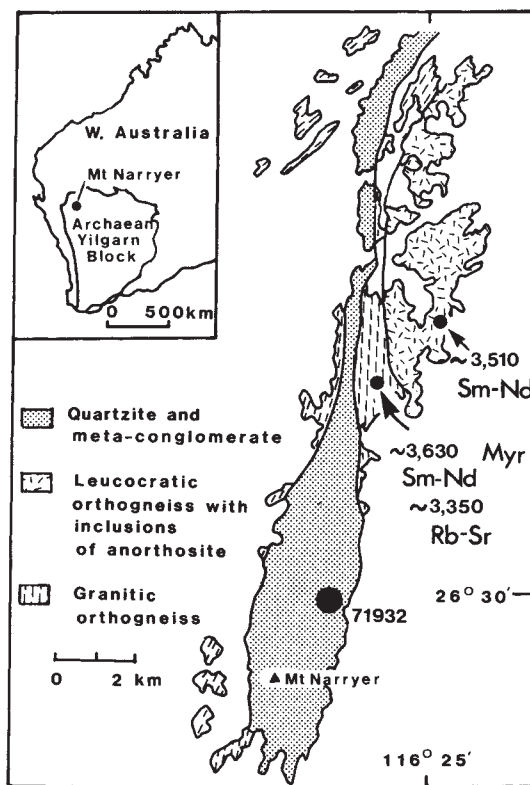


Fig. 1 Geological sketch map (slightly simplified, after ref. 1) with sample locality and previous model Sm/Nd and Rb/Sr ages² for gneisses adjacent to the quartzite.

data in order of decreasing U/Pb ages. Such high-resolution analyses have been shown to be very useful for deciphering complex zircon populations in igneous and metamorphic rocks (for example, refs 3, 4) and elucidating the age spectrum (provenance ages) of detrital zircons^{5,6}. The analysis of fragments of single grains enables us to determine intra-grain ages⁷. Figure 1 shows the location of the sample and indicates previous model Sm/Nd ages² from the adjacent gneisses as well as a Rb/Sr age², interpreted as dating high-grade metamorphism in the gneisses.

* Present address: Royal Ontario Museum, Mineralogy and Geology Department, 100 Queen's Park, Toronto, Ontario, Canada M5S 2C6.

Table 1 U/Pb analytical results of single zircon grains and single fragments, Mt Narryer metaquartzite, Western Australia

Grain No.	Type	U (pmol)*	Pb _{rad} † (pmol)	Pb _c ‡ (pmol)	²⁰⁶ Pb/ ²⁰⁴ Pb (measured)	Radiogenic Pb (atomic%)			Atomic ratios			Apparent ages (Myr)		
						²⁰⁶ Pb	²⁰⁷ Pb	²⁰⁸ Pb	²⁰⁶ Pb/ ²³⁸ U	²⁰⁷ Pb/ ²³⁵ U	²⁰⁷ Pb/ ²⁰⁶ Pb	²⁰⁶ Pb/ ²³⁸ U	²⁰⁷ Pb/ ²³⁵ U	²⁰⁷ Pb/ ²⁰⁶ Pb
1	A	9.108	8.612	0.275	929	70.77	23.05	6.18	0.6740	30.27	0.3257	3,321	3,495	3,597
2	A	12.15	10.92	—§	3,370	71.27	22.44	6.29	0.6450	28.01	0.3150	3,209	3,419	3,546
3	A	6.818	5.848	—	1,400	68.22	21.96	9.82	0.5888	26.14	0.3220	2,984	3,352	3,579
4	B	9.990	8.295	0.203	1,134	73.55	22.05	4.40	0.6156	25.42	0.2995	3,092	3,324	3,468
5	B	11.57	9.686	—	2,542	74.46	21.42	4.12	0.6279	24.88	0.2874	3,141	3,304	3,404
6	B	3.634	3.029	—	912	71.25	20.84	7.91	0.5984	24.13	0.2925	3,023	3,274	3,431
7	C	8.171	7.093	—	1,761	69.13	19.93	10.94	0.6044	24.07	0.2888	3,048	3,271	3,411
8	C	6.390	5.134	—	2,762	74.10	21.28	4.62	0.5985	23.70	0.2872	3,024	3,256	3,403
9	A	15.02	10.98	0.212	1,466	75.46	21.69	2.85	0.5557	22.02	0.2874	2,849	3,185	3,404
10	B	13.65	10.65	0.555	701	74.85	20.28	4.87	0.5884	21.97	0.2708	2,983	3,182	3,311
11	A	14.37	10.69	0.401	887	73.72	20.59	5.69	0.5522	21.26	0.2792	2,834	3,151	3,358
12	C	2.096	1.585	—	493	75.16	19.68	5.16	0.5716	20.71	0.2628	2,914	3,125	3,265
13	C	13.28	8.703	0.102	1,784	74.95	21.15	3.90	0.4946	19.25	0.2823	2,591	3,054	3,376
14	A	2.151	1.556	0.314	170	74.57	19.05	6.38	0.5439	19.12	0.2550	2,800	3,048	3,217
15	B	3.109	2.177	0.135	395	75.87	19.43	4.70	0.5343	18.86	0.2560	2,760	3,035	3,222
16	A	4.844	3.470	—	993	73.33	17.86	8.81	0.5289	17.76	0.2435	2,737	2,977	3,143
17	A	8.642	5.702	0.102	1,278	76.63	19.27	4.10	0.5092	17.66	0.2515	2,653	2,971	3,195
18	A	8.289	5.651	0.101	1,152	72.83	18.86	8.31	0.4998	17.86	0.2592	2,613	2,982	3,242
19	D	31.15	19.52	0.700	1,061	76.84	19.87	3.29	0.4822	17.19	0.2586	2,537	2,945	3,238
20	A	6.306	3.942	—	1,158	76.44	19.59	3.97	0.4829	17.02	0.2556	2,540	2,936	3,220
21	D	33.24	19.77	1.419	600	77.98	17.38	4.64	0.4671	14.35	0.2228	2,471	2,773	3,001
22	D	15.41	9.183	1.009	358	72.87	16.23	10.90	0.4372	13.49	0.2238	2,338	2,715	3,008
23	A	23.09	13.19	2.167	265	75.06	16.10	8.84	0.4316	12.77	0.2146	2,313	2,663	2,941
24	A	21.67	12.83	2.794	192	69.99	15.26	14.75	0.4176	12.56	0.2181	2,250	2,647	2,967
25	D	56.38	29.88	3.770	352	75.41	16.47	-8.12	0.4023	12.11	0.2183	2,180	2,613	2,968
26	D	24.36	11.72	1.052	463	77.32	16.34	6.34	0.3747	10.92	0.2114	2,051	2,516	2,917
27	C	51.19	22.77	0.989	992	82.00	15.72	2.28	0.3675	9.711	0.1916	2,018	2,408	2,756
Single fragment analyses														
28	A I	1.248	1.144	0.135	210	71.26	21.61	7.13	0.6582	27.49	0.3029	3,260	3,401	3,485
	II	4.663	4.207	0.174	627	72.23	21.91	5.86	0.6550	27.41	0.3035	3,248	3,398	3,488
	III	3.277	2.923	0.738	157	71.68	21.49	6.83	0.6442	26.59	0.2994	3,206	3,368	3,467
29	D I	4.197	3.150	0.125	581	71.71	20.71	7.58	0.5426	21.60	0.2887	2,794	3,166	3,412
	II	1.622	1.173	0.232	154	73.39	20.53	6.08	0.5358	20.53	0.2779	2,766	3,117	3,351
	III	15.71	10.39	0.347	1,007	75.67	20.45	3.88	0.5034	18.76	0.2703	2,628	3,030	3,308
30	C I	7.474	5.487	1.255	182	72.76	20.46	6.78	0.5368	20.84	0.2816	2,770	3,131	3,372
	II	9.814	6.170	0.174	915	72.75	20.71	6.53	0.4606	18.08	0.2847	2,442	2,994	3,389
31	A I	8.701	5.744	0.251	676	74.07	19.49	6.44	0.4924	17.87	0.2632	2,581	2,983	3,267
	II	7.566	4.755	0.280	512	73.04	18.79	8.17	0.4624	16.42	0.2575	2,450	2,902	3,232
32	B I	13.28	7.342	1.684	188	73.68	14.39	11.93	0.4105	11.04	0.1951	2,217	2,527	2,786
	II	55.27	26.55	3.118	395	79.57	14.80	5.64	0.3848	9.869	0.1860	2,099	2,423	2,707

Types A, B, C, D correspond to the shape characteristics given in Fig. 2b. The constants used are: $1.55125 \times 10^{-10} \text{ yr}^{-1}$ for $\lambda^{238}\text{U}$ (ref. 15), $9.8485 \times 10^{-10} \text{ yr}^{-1}$ for $\lambda^{235}\text{U}$ (ref. 15) and 137.88 for $^{238}\text{U}/^{235}\text{U}$ (ref. 16). Zircons were dissolved in Teflon bombs¹⁷ in 50% HF during ~100 h at 220 °C. Element separation and purification were performed as described earlier¹⁸. The isotope ratios and amounts were determined by isotope dilution analyses in a Thomson THN 206 mass spectrometer using a $^{205}\text{Pb}/^{235}\text{U}$ mixed tracer. Both elements were loaded with H_3PO_4 and silica gel on Re single filaments. The isotope composition of blank Pb was determined to be 17.84 ± 0.10 (2σ error) for $^{206}\text{Pb}/^{204}\text{Pb}$, 15.53 ± 0.08 for $^{207}\text{Pb}/^{204}\text{Pb}$ and 37.41 ± 0.21 for $^{208}\text{Pb}/^{204}\text{Pb}$. The total Pb blank was $30 \pm 15 \times 10^{-12} \text{ g}$ ($5\text{--}10 \times 10^{-12} \text{ g}$ from the chemical procedure and $10\text{--}35 \times 10^{-12} \text{ g}$ extracted from the Teflon bombs). The total blank for U was $<10 \times 10^{-12} \text{ g}$ with an average of $\sim 5 \times 10^{-12} \text{ g}$. The maximum errors on the isotope ratios are 0.5% for $^{206}\text{Pb}/^{238}\text{U}$ and $^{207}\text{Pb}/^{206}\text{Pb}$ and 0.7% for $^{207}\text{Pb}/^{235}\text{U}$. These errors take into account (1) the precision of the mass spectrometric measurements, (2) the uncertainties in the isotope composition and amount of blank Pb, (3) the correction for ^{205}Tl interference, (4) the uncertainty of mass-discrimination corrections and (5) the accuracy of tracer calibration.

* Absolute amount of U in the sample, corrected for $5 \times 10^{-12} \text{ g}$ U blank.

† Absolute amount of radiogenic Pb in the sample, corrected for $30 \times 10^{-12} \text{ g}$ blank Pb and, where detectable, corrected also for incorporated common Pb (Pb_c) with the following isotope composition: 11.83 for $^{206}\text{Pb}/^{204}\text{Pb}$, 13.87 for $^{207}\text{Pb}/^{204}\text{Pb}$ and 31.67 for $^{208}\text{Pb}/^{204}\text{Pb}$. This isotope composition corresponds to a 3,630 Myr BP Pb that evolved from 4,550 Myr BP with a $^{238}\text{U}/^{204}\text{Pb}$ of 9.3, a $^{232}\text{Th}/^{204}\text{Pb}$ of 39.1 and a primordial isotope composition identical to the Canyon Diablo meteorite¹⁹; further comments on Pb_c in zircon are given in the text.

‡ Absolute amount of Pb_c in the sample.

§ No incorporated Pb_c detectable in the sample (total amount of Pb_c in the analyses was $<45 \times 10^{-12} \text{ g}$).

Figure 2a is a summary Concordia diagram⁸ of single-grain and fragment analyses together with the ion-probe data for the apparently 4,100–4,200-Myr BP grains and Fig. 2b shows a more detailed Concordia diagram with a broad grouping of the different zircon types and a distinction between single-grain and individual fragment data. Additionally, Fig. 2b indicates the upper intercept ages of the regression lines through distinct fragments of single grains (intra-grain ages).

All zircons analysed grain-by-grain were chosen by their habit, colour, degree of transparency, size and particular features such as pronounced core-rim zoning. Those from which the fragments were broken were selected to cover the whole range of different grain types. Because Froude *et al.*¹ indicated that the apparently 4,100–4,200-Myr zircons are distinguished by their peculiar angular form and the absence of inclusions, we first investigated perfectly euhedral grains with sharp edges, high transparency (to avoid inclusions) and uncorroded surfaces. However, we mention that the grain illustrated for reference in a photomicrograph in Froude *et al.*¹ resembles more a simple piece of broken zircon than a particular angular type. Nevertheless, we analysed

in a first group of zircons (see Fig. 2b) the series of grains that most closely fulfilled the above conditions, followed by a second group (Fig. 2b) similar to the first but with much lower degrees of transparency. In a third group we investigated transparent and translucent grains with strongly rounded edges, some of which had significant surface corrosion features. As a last group with distinct grain type, a series of particularly milky translucent zircons was analysed, most of them exhibiting clearly visible zoning between the inner and outer parts of the crystal. This latter feature suggests that some of the zircons grow in two stages.

Despite the wide range of grain characteristics, all zircons yield similarly discordant age patterns (Fig. 2a, b) with minimum ages ranging from 2,750 to 3,600 Myr (Table 1), plotting in a data field that intersects the Concordia curve along a sector from ~3,400 to 3,700 Myr. This suggests that the primary ages of the zircons lie in this interval, but because none of the different groups of zircons defines a simple linear array, we cannot distinguish cogenetic zircon populations and, thus, no precise primary ages can be determined.

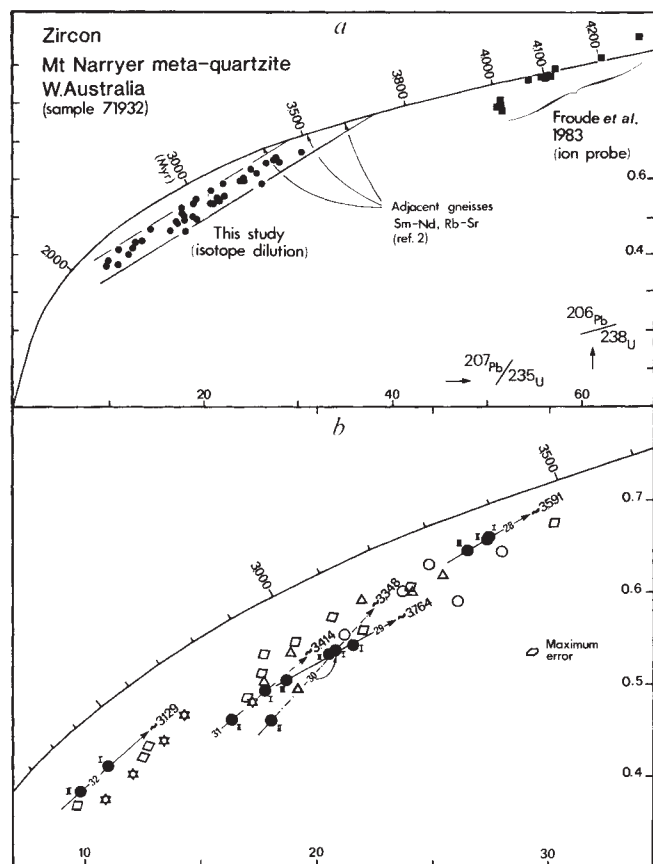


Fig. 2 Concordia diagram of the U/Pb analytical results. *a*, Summary diagram of single-grain and fragment analyses compared with ion-probe data¹ for the apparently 4,100–4,200-Myr BP zircons; *b*, detailed diagram showing distinctions between different grain types and single-grain versus fragment analyses. Single zircon grains: $\square$, Sharp-edged transparent grains; $\triangle$, sharp-edged translucent grains; $\circ$, strongly edge-rounded transparent and translucent grains; $\star$, milky translucent grains with core-rim zoning; $\bullet$, single fragments (from grains 28–32).

A few observations concerning the grain-by-grain analysis may be emphasized: the particularly milky translucent grains yield younger ages than most of the other grain types, plotting in the lowest segment of the discordancy field (Fig. 2*b*). Moreover, these milky grains are, together with some other highly discordant grains, characterized by especially large amounts of incorporated common lead (Pb_c , Table 1), contributing $\leq 15\%$ of total lead in the crystal. This positive correlation between Pb_c content and degree of discordancy supports an earlier conclusion⁷ that loss of radiogenic lead may occur by exchange with Pb_c from intergranular phases. The presence of such incorporated Pb_c , also observed to a smaller extent in less discordant grains, requires a correction which necessarily affects the apparent ages of all Pb_c -bearing grains. We performed this correction by assuming the incorporated Pb_c to be on average $\sim 3,630$ Myr old, as suggested by the minimum ages of Pb_c -free grains ranging up to $\sim 3,600$ Myr, and supported by the model Sm/Nd ages² for the adjacent gneisses, the possible source rocks of the quartzite. In the case of sediments, the isotope composition of this incorporated common lead cannot be determined directly because of the lack of Pb-rich/U-deficient cogenetic minerals such as K-feldspar. Therefore, the above assumption is necessary. However, a correction with a significantly younger lead (for example, a 3,350-Myr-old Pb_c) (probable age of metamorphism in the quartzite) would only cause age differences lying within the maximum error of our analytical technique (Table 1, Fig. 2*b*). A correction with a 4,100-Myr-old Pb would influence a few data by 2%, whereas most of the single-grain data would be changed by $<1\%$. This shows that variations in the isotope composition of incorporated Pb_c , relative to the

3,300–4,100-Myr age range, modify the data only to a small extent; even a correction with present-day Pb_c (for example, blank Pb_c , Table 1) would not shift any of the data to ages older than 3,700 Myr.

Our results from the single-grain analyses confirm neither the existence of 4,100–4,200-Myr-old zircons in this sample nor the particular discordancy patterns observed on the ion probe¹. The discordancy arrays of the ion-probe data all point to the origin of the Concordia curve (zero ages), suggesting recent Pb loss (or U gain) of these zircons (see Fig. 2*a* and ref. 1). Here, we do not show any evidence in favour of such recent perturbation of the U/Pb systems; the lower extrapolation of our discordancy field for all single-grain data intersects the Concordia curve along a sector from $\sim 1,400$ to 2,000 Myr, substantiating Pb to be ancient. Moreover, none of our grains yield a result that plots above the Concordia curve as reported for the ion-probe analyses for both apparently 4,100-Myr-old and significantly younger zircons (see Fig. 2*a* and ref. 1).

Because of the discrepancy between the isotope dilution and the ion-probe data, we extended our investigation to analyses of individual fragments of single grains, for two main purposes: (1) to assess whether distinct parts of grains preserved a record of zircon growth before 4,000 Myr BP and (2) to obtain intra-grain ages, which approximate primary zircon ages better than those of arrays of single, and not necessarily cogenetic, zircon grains. Hence, we broke up five zircons representative of the different grain types observed in the quartzite (Fig. 2*b*).

A typical grain of the first group (grain 28), a sharp-edged transparent zircon, was broken into three fragments (I, II, III, Fig. 2*b*, Table 1), where I represents the middle part of this long prismatic grain and II and III the two prismatic ends. The regression line⁹ through the fragments (calculated using the individual errors of the fragment analyses) intercepts the Concordia curve at $\sim 3,591$ Myr, which is in agreement with the model Sm/Nd age for the possible source rocks of the quartzite. Three fragments of another grain (grain 29), a sharp-edged but translucent zircon (second group in Fig. 2*b*), define an intercept age of $\sim 3,764$ Myr. As for grain 28, the central part of the long prismatic zircon (fragment I) is the oldest, suggesting that loss of radiogenic Pb is favoured in the pyramidal ends of the grain either because of crystallographic differences between middle and outer parts or simply because of the higher surface/volume ratio of the pyramidal ends. A third grain (grain 30), a particularly milky translucent and core-rim-zoned zircon (fourth group in Fig. 2*b*), has been analysed in two fragments (I and II) where I consisted mainly of rim material, whereas II was dominated by material from the core of the grain. The two fragments define a line with an intercept age of $\sim 3,348$ Myr, which is identical to the probable age of high-grade metamorphism, in the adjacent gneisses². To account for this age coincidence as well as the fact that grain 30 was core-rim zoned, it seems plausible that the intercept age dates zircon overgrowth (that is, new growth) during regional metamorphism, which is recognized clearly in the quartzite¹⁰. This interpretation seems to be supported by the discordancy trend for single grains of milky zoned zircons (Fig. 2*b*), which points to significantly younger ages than the arrays of most other zircons. A further zircon of the first group (grain 31), a transparent sharp-edged homogeneous grain, broken in two approximately equal fragments, yields an intercept age of $\sim 3,414$ Myr, suggesting that some of the smaller grains represent entirely new-grown zircons during high-grade metamorphism. A final zircon, broken in two fragments (grain 32), was a peculiar green milky translucent grain yielding an intercept age of $\sim 3,129$ Myr. This age suggests the existence of a second younger metamorphic event in the quartzite. However, the line through the two fragments is too poorly defined to define the intercept age with any confidence and, moreover, no such event is known in the area.

Thus, individual fragments from single grains representing the different crystal types in the quartzite also do not yield any evidence for zircons older than 3,800 Myr. Furthermore, the discordancy trends of the $\sim 3,600$ –3,800-Myr BP fragments

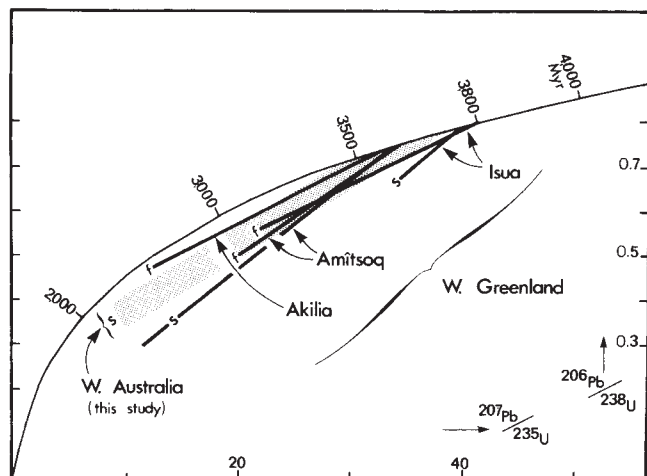


Fig. 3 Summary Concordia diagram including the Australian Mt Narryer quartzite zircons analysed here and, for comparison, zircons from Western Greenland, specifically from the Isua supracrustals^{12,13}, the Amîtsoq gneisses¹²⁻¹⁴ and the Akilia association¹³. All ages have been calculated using the new uranium decay constants¹⁵; f, fractions; s, single grains.

(grains 28 and 29) are identical to those for the single zircon grains, again contrary to the discordancy patterns obtained on the ion probe¹. The question remains as to whether the lack of 4,100–4,200-Myr BP zircons found here is an artefact of grain selection, that is, that we have not chosen the correct crystals. This possibility cannot be completely eliminated. Furthermore, there is an independent argument to be considered: the single grains and individual fragments analysed here yield clearly discordant ages, proving that all zircons were open systems for Pb either during a distinct period or continuously. This contrasts strongly with the apparently concordant ages of the 4,100–4,200-Myr BP zircons identified by the ion-probe technique. Such closed-system behaviour of the 4,100–4,200 Myr BP zircons, which would necessarily have a higher degree of radioactive damage than the younger ones, is difficult to explain. Closed-system behaviour would be particularly surprising given that the quartzite from which the zircons have been extracted suffered high-grade metamorphism probably ~3,400 Myr ago; in general, even low-grade metamorphism, for example, under lower greenschist facies conditions, produces significant Pb loss in detrital zircons (for example, ref. 11).

In conclusion, we emphasize that the complete data set for single grains and fragments can be explained by zircon growth between 3,300 and 3,800 Myr BP. The upper intercept ages of both the overall discordancy field and two of the intra-grain arrays point to primary ages of the order of 3,600–3,800 Myr. Because the quartzite has been affected by high-grade metamorphism, probably ~3,400 Myr ago, these upper intercept ages may be slightly too young relative to the real crystallization ages of zircon, because of an early Pb loss during metamorphism. From this point of view, the possibility that parts of the Australian continent are older than 3,800 Myr cannot be excluded completely. However, the present study does not produce any evidence in support of this idea. On the other hand, the zircon ages reported here show the Australian continent to contain crustal rocks 3,600–3,800 Myr in age, analogous to those in Greenland. This similarity is illustrated in Fig. 3, which is a summary Concordia diagram including all reliable zircon ages from Western Greenland (Isua, Amîtsoq and Akilia; refs 12–14) together with our data from the Mt Narryer quartzite. The overlap between possible primary ages and discordancy arrays of the Australian and Greenland zircons is consistent with the hypothesis that both cratons in the present-day Northern and Southern Hemisphere developed from a continental nucleus that grew ~3,600–3,800 Myr ago.

We thank Dr W. Compston for providing the zircons in which the 4,100–4,200-Myr-old grains were identified (sample GWSA

71932) and S. H. Richardson and G. Manhès for critical reading of the manuscript. This is Institut de Physique du Globe contribution 813.

Received 23 October 1984; accepted 12 February 1985.

1. Froude, D. O. *et al.* *Nature* **304**, 616–618 (1983).
2. De Laeter, J. R. *et al.* *Nature* **292**, 322–324 (1981).
3. Lancelot, J., Vitrac, A. & Allègre, C. J. *Earth planet. Sci. Lett.* **29**, 357–366 (1976).
4. Schärer, U. & Allègre, C. J. *Earth planet. Sci. Lett.* **63**, 423–432 (1983).
5. Gaudette, H., Vitrac-Michard, A. & Allègre, C. J. *Earth planet. Sci. Lett.* **54**, 248–260 (1981).
6. Schärer, U. & Allègre, C. J. *Can. J. Earth Sci.* **19**, 1910–1918 (1982).
7. Schärer, U. & Allègre, C. J. *Nature* **295**, 585–587 (1982).
8. Wetherill, G. *Trans. Am. geophys. Un.* **37**, 320–326 (1956).
9. Williamson, J. H. *Can. J. Phys.* **46**, 1845–1847 (1968).
10. Williams, I. R., Walker, I. W., Hocking, R. M. & Williams, S. J. *Geol. Surv. Western Australia* **5** (1980).
11. Schärer, U. *Earth planet. Sci. Lett.* **49**, 205–218 (1980).
12. Michard-Vitrac, A., Lancelot, J. & Allègre, C. J. *Earth planet. Sci. Lett.* **35**, 449–453 (1977).
13. Baadsgaard, H. *et al.* *Earth planet. Sci. Lett.* **68**, 221–228 (1984).
14. Baadsgaard, H. *Earth planet. Sci. Lett.* **19**, 22–28 (1973).
15. Jaffey, A. H., Flynn, R. F., Glendenin, L. E., Bentley, W. C. & Essling, A. M. *Phys. Rev.* **C4**, 1889–1906 (1971).
16. Shields, W. R. *US natn. Bur. Stand. Meet. Rep.* **8**, 1–37 (1960).
17. Krogh, T. E. *Geochim. cosmochim. Acta* **37**, 488–494 (1973).
18. Manhès, G., Minster, J. F. & Allègre, C. J. *Earth planet. Sci. Lett.* **39**, 14–24 (1978).
19. Tatsumoto, M., Knight, R. J. & Allègre, C. J. *Science* **180**, 1279–1283 (1973).

Halomethane from halide ion—a highly efficient fungal conversion of environmental significance

David B. Harper

Department of Agricultural and Food Chemistry, The Queen's University of Belfast, and Agricultural and Food Chemistry Research Division, Department of Agriculture for Northern Ireland, Newforge Lane, Belfast BT9 5PX, UK

Globally, at least 5 Mt yr⁻¹ of chloromethane (CH₃Cl) must originate from natural sources, according to estimates based on environmental concentrations, man-made emissions of 26 kt yr⁻¹ being insignificant in comparison^{1–3}. Although CH₃Cl is a natural product of several fungal species^{4,5}, no quantitative investigations of its biological production have been reported. We have measured CH₃Cl production by a common wood-rotting fungus, and report that on glucose-based media, the chloride ion (Cl⁻) at concentrations <4 mM was methylated with >90% efficiency. The pattern of CH₃Cl biogenesis was typical of a secondary metabolite and paralleled loss of Cl⁻ from the medium. With cellulosic substrates, CH₃Cl yields ranged between 80 and 95% at Cl⁻ concentrations as high as 25 mM, although production extended over a longer period. Bromo- and iodomethane were formed in high yield from the corresponding halide ions. Such high-efficiency methylation of the halide ion indicates that fungi could make a substantial contribution to the atmospheric CH₃Cl burden.

Assessments of CH₃Cl production were made during growth of *Phellinus pomaceus* (Pers.) syn. *Fomes pomaceus* (Forest Products Research Laboratory no. 33A), a white-rot fungus of worldwide distribution reported to synthesize CH₃Cl (ref. 4). When the fungus was grown with glucose as the main carbon source, CH₃Cl production was clearly confined (Fig. 1) to the period after exponential growth and before autolysis, a pattern characteristic of secondary metabolites, the formation of which is usually retarded until the trophophase approaches completion^{6,7}. Whereas fungal growth continued for ~10 days, CH₃Cl biosynthesis extended for another 8 days. Chloride uptake from the medium closely paralleled CH₃Cl biogenesis, showing that there was little accumulation of Cl⁻ in the mycelium before CH₃Cl release, a conclusion confirmed by determination of Cl⁻ content in the mycelium after ashing at various stages of growth. Chloromethane yield, expressed as a

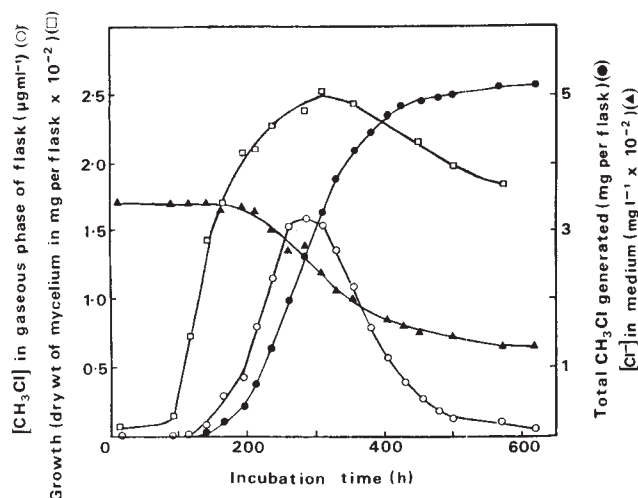


Fig. 1 Growth of *P. pomaceus* on glucose-based medium containing 9.6 mM NaCl in relation to CH_3Cl production and chloride loss from the medium. Results are expressed as follows: $\square$, growth measured as mycelial dry wt; $\circ$, CH_3Cl concentration in gaseous phase in the culture flask; $\bullet$, total CH_3Cl generated per culture flask; $\blacktriangle$, chloride ion concentration in the culture medium.

Methods. Cultures of *P. pomaceus* were grown on a solid medium pH 6.5 (20 ml) containing (g l^{-1}): glucose, 30; mycological peptone, 5; agarose '10' (BDH), 10; NaCl, 0.5 in conical flasks (25×250 ml) at 25°C . As a precaution against bacterial contamination, chloramphenicol (25 mg l^{-1}) was also added to the medium, experiments having indicated that the presence of the compound did not affect production of CH_3Cl or chloride levels in the medium during growth. The inoculum, which was spread uniformly over the surface of the medium, was a mycelial suspension (1 ml) prepared by homogenizing mycelium (600 mg wet wt) in sterile distilled water (20 ml) for 30 s with an Ultra Turrax homogenizer. The mycelium used for inoculation was grown on 5% malt extract agar containing not more than 1 mM chloride. Each conical flask was provided with a polytetrafluoroethylene (PTFE)-coated rubber bung through which passed a glass tube (10 cm $\times$ 0.8 cm internal diameter, i.d.) packed uniformly with cotton wool and fitted so that the lower end was flush with the lower surface of the stopper. The bung was also equipped with a sampling port comprising a glass tube of 1 mm i.d. projecting 0.5 cm into the interior of the flask and flared at the upper end protruding from the stopper so as to accommodate a small PTFE-coated rubber bung. CH_3Cl production during growth was assessed by a modification of a headspace technique described previously¹⁴. Initially, to calibrate the system, the rate of loss of gaseous halomethane from flasks equipped with outlet tubes as above was measured by introducing aqueous solutions of halomethane (20 ml) and determining headspace concentrations (by gas chromatography with a flame ionization detector) of aliquots (2 ml) withdrawn through the sampling port at intervals over 8 days. As diffusion of halomethane from the liquid to the gaseous phase within the flask was rapid compared with the rate of loss from the flask, the system could be regarded as single-phase in subsequent calculations. The loss rate obeyed first-order kinetics, the half-lives of chloro-, bromo- and iodomethanes being 13.8, 19.6 and 26.2 h respectively. Therefore, the total amount of halomethane generated by a fungal culture at any time during growth could be calculated from empirical measurements, at regular intervals, of the halomethane concentration in the flask headspace and the half-life of the halomethane using a computer integrating program¹⁴. Growth was measured as the dry wt of washed mycelium obtained by filtration of the agarose medium after solution in hot water. Chloride ion concentrations were determined on the diluted medium with a Corning Eel 921 chloride meter.

percentage of the Cl^- originally present in the medium, varied between 90 and 100% for Cl^- concentrations <4 mM, falling to 6% at 170 mM Cl^- . Between 1 and 170 mM Cl^- , total CH_3Cl production was linearly related to the logarithm of Cl^- concentration, although slight deviation occurred at the higher end of the concentration range.

As the fruiting bodies of *P. pomaceus* are typically found on trees of the Rosaceae, especially *Prunus*, CH_3Cl yields were

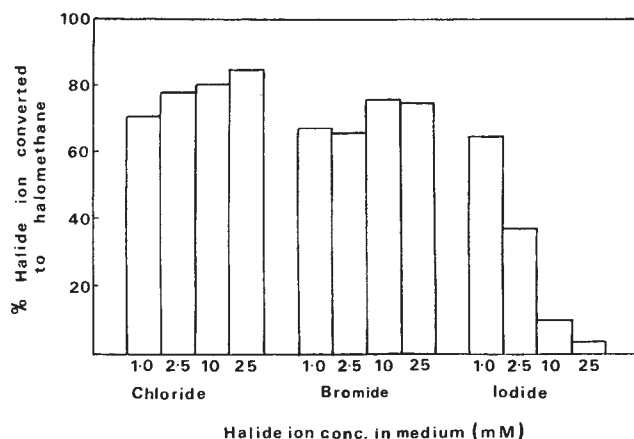


Fig. 2 Effect of halide ion concentration on yield of corresponding halomethane for *P. pomaceus* grown on cellulose as main carbon source.

Methods. Cultures of *P. pomaceus* were grown in conical flasks (in the conditions described in Fig. 1 legend) on filter paper (Whatman No. 540) (6.9 g) augmented with a chloride-free mineral salts medium of pH 6.5 (20 ml) containing pure L-amino acids (3.6 g l^{-1}) in similar proportions to those in mycological peptone and thiamine mononitrate (1.2 mg l^{-1}). The medium was supplemented with 1–25 mM NaCl, NaBr or NaI, as indicated. The results shown are the mean of duplicate treatments.

assessed when the fungus was cultured with supplemental NaCl (20 mM) on sawdust derived from the wood of *Prunus domestica*, the common plum. Chloromethane production with both wood and other cellulosic substrates (such as filter paper, microcrystalline cellulose, cotton wool and newspaper) as main carbon sources extended over long periods, from 4 weeks (cotton wool) to 6 months (wood) but yields of CH_3Cl were much higher than with glucose-based medium at similar chloride concentration, ranging from 86% (cotton wool) to 95% (newspaper).

Figure 2 illustrates the effect of different concentrations of halide ion on halomethane yield, with cellulose (filter paper) as the substrate. Both bromide and iodide acted as acceptors for the methylating system. In contrast, with glucose-based media, yields of both chloro- and bromomethane did not decline with increasing halide ion concentration; indeed, a slight trend in the reverse direction was discernible with Cl^- . The suppression of secondary metabolite production by glucose and other readily utilizable carbon sources has been noted in many studies on secondary metabolism⁷. Iodomethane production decreased with increasing iodide concentration but this is probably attributable to the manifestly detrimental effects of this ion on growth. When equimolar concentrations (5 mM) of chloride, bromide and iodide were present, the preferred substrate was iodide, the respective halomethanes being formed in the proportion 1:4:19.

The magnitude of CH_3Cl production in nature by fungi is difficult to estimate. Chloromethane has been identified as a volatile product of six *Fomes* species⁴ and of *Agaricus bisporus*⁵, the common mushroom. Future studies may reveal the presence of this biosynthetic ability in many more cellulose-degrading species growing on wood, litter and soil worldwide. The high affinity of the methylating system for Cl^- ensures that even when environmental Cl^- concentrations are low, appreciable quantities of CH_3Cl are generated. Thus, despite a concentration of only 34 mg kg^{-1} in *P. domestica* wood, yields of CH_3Cl when *P. pomaceus* was grown on the unsupplemented substrate were between 40 and 50%. However, wood-rotting fungi are not confined to such low-chloride environments, being capable of mycelial growth on many plant substrates under a broad range of environmental conditions. The normal Cl^- content of plant material varies from 200 to 10,000 mg per kg dry wt^{8,9} and many terrestrial soils exhibit appreciable salinity. Growth of fungi possessing such a high-affinity methylating system in these

circumstances could lead to production of substantial quantities of CH_3Cl and must therefore be considered alongside other postulated sources of atmospheric CH_3Cl such as combustion of vegetation¹⁰ and reaction of iodomethane with Cl^- in sea water¹¹. That such a biological source may be a major contributor to atmospheric CH_3Cl levels is of considerable interest in view of the importance ascribed to CH_3Cl and other man-made halocarbons in controlling the levels of stratospheric ozone^{11,12}. Quite apart from the environmental significance, the enzymatic incorporation of halide ion apparently directly into a one-carbon

compound deserves attention as it must involve a mechanism quite distinct from the haloperoxidase-mediated halide ion incorporation previously regarded as the major route to organohalogen compounds in nature¹³. The biotechnological potential of this novel system warrants investigation; indeed, a role for the fungal gene involved can be envisaged in the development of halotolerant crop plants, although environmental considerations may limit such an application.

I thank J. T. Kennedy for technical assistance and J. A. Buswell, P. Robinson and D. A. Seaby for helpful discussions.

Received 17 January; accepted 4 March 1985.

1. Lovelock, J. E. *Nature* **256**, 193–194 (1975).
2. Singh, H. B., Salas, L. J., Shigeishi, H. & Scribner, E. *Science* **203**, 899–903 (1979).
3. Edwards, P. R., Campbell, I. & Milne, G. S. *Chem. Ind.* **1982**, 619–622 (1982).
4. Cowan, M. I. *et al. Trans. Br. mycol. Soc.* **60**, 347–351 (1973).
5. Turner, E. M., Wright, M., Ward, T., Osborne, D. J. & Self, R. J. *gen. Microbiol.* **91**, 167–176 (1975).
6. Weinberg, E. D. *Adv. micro. Physiol.* **4**, 1–43 (1970).

7. Drew, S. W. & Wallis, D. A. in *Secondary Metabolism and Differentiation in Fungi* (eds Bennett, J. W. & Ciegler, A.) 35–51 (Dekker, New York, 1983).
8. Bethge, P. O. & Troeng, T. *Svensk Papp-Tidn.* **62**, 598–601 (1959).
9. Long, C. in *Biochemists' Handbook*, 1040–1041 (E. & F. N. Spon Ltd, London, 1961).
10. Palmer, T. Y. *Nature* **263**, 44–46 (1976).
11. Zafriou, O. C. *J. mar. Res.* **33**, 75–81 (1975).
12. Edwards, P. R., Campbell, I. & Milne, G. S. *Chem. Ind.* **1982**, 574–578 (1982).
13. Neidleman, S. L. & Geigert, J. *Biochem. Soc. Symp.* **48**, 39–52 (1983).
14. Harper, D. B. & Nelson, J. J. *gen. Microbiol.* **128**, 1667–1678 (1982).

Use of numbers by a chimpanzee

Tetsuro Matsuzawa

Department of Psychology, Primate Research Institute, Kyoto University, Inuyama, Aichi 484, Japan

Recent studies have examined linguistic abilities in apes^{1–6}. However, although human mathematical abilities seem to be derived from the same foundation as those in language, we have little evidence for mathematical abilities in apes (but for exceptions see refs 7–10). In the present study, a 5-yr-old female chimpanzee (*Pan troglodytes*), 'Ai', was trained to use Arabic numerals to name the number of items in a display. Ai mastered numerical naming from one to six and was able to name the number, colour and object of 300 types of samples. Although no particular sequence of describing samples was required, the chimpanzee favoured two sequences (colour/object/number and object/colour/number). The present study demonstrates that the chimpanzee was able to describe the three attributes of the sample items and spontaneously organized the 'word order'.

Before the present study^{11,12,14}, Ai had been trained to name 14 objects and 11 colours by choosing among the set of symbols shown in Fig. 1. For the first experiment, five objects and five colours were selected from Ai's 'vocabulary' to form groups of sample items to be identified. The experimenter presented the sample items (for example, three red pencils) in the display

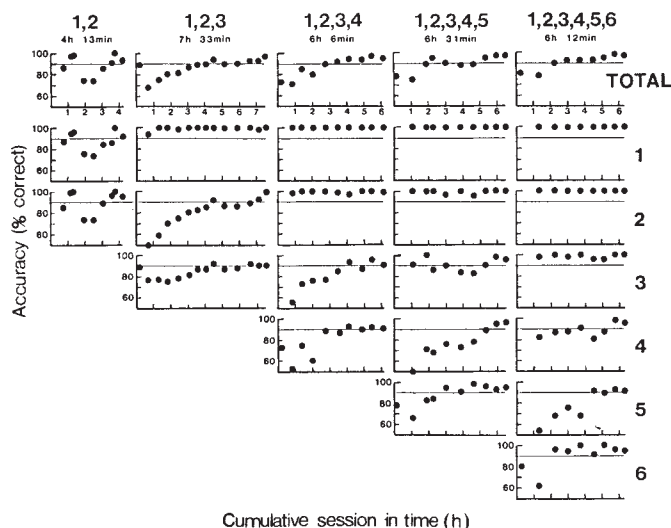


Fig. 2 Acquisition of number names for the training item, red pencils. The top row represents overall accuracy, the other rows accuracy for individual numbers. Each dot shows the performance in each session, with percentage correct plotted against the time elapsed since the beginning of training on each number set. In the first session of each stage, except for training on the number names 1 and 2, the chimpanzee received intensive training limited to the newly introduced number.

Fig. 1 Symbols used in these studies. In the symbolic matching-to-sample task, the chimpanzee was trained to identify the sample item that appeared in the display window (24 × 17 × 24 cm) by pressing an illuminated key (2 × 2.5 cm) showing the assigned symbol. These symbols could appear in various positions in a 5 × 6 key matrix on a console located below the display window. This console was interfaced with a PDP11/V03 minicomputer. The set of operative-illuminated keys was changed from trial to trial and the positions of the keys were also changed from session to session so that the symbol, rather than the position of the correct key, had to be learned. The trial ended when the chimpanzee pressed a blank key to the right of the matrix on the console. In numerical naming, two methods of displaying the sample items were used to minimize the possibility that the chimpanzee would associate each numeral with a particular display pattern. On some trials the samples were laid out in a row. The position and the spatial density of items were changed within the area of the window in each trial so that area covered was not correlated with number of items. On other trials the items were held in one hand in front of the display window. The orientation and the angular separation of items was varied from trial to trial. Depending on type of object, either or both methods were used. In the first experiment, five objects (pencil, paper, brick, spoon and tooth-brush) of five colours (red, green, blue, yellow and black) were used to form groups of items to be identified.

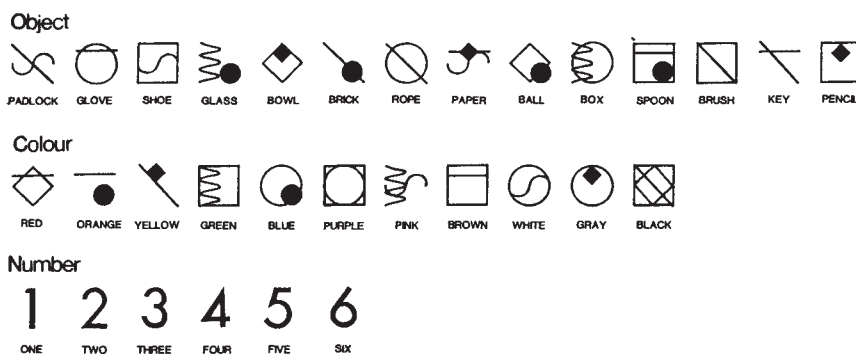
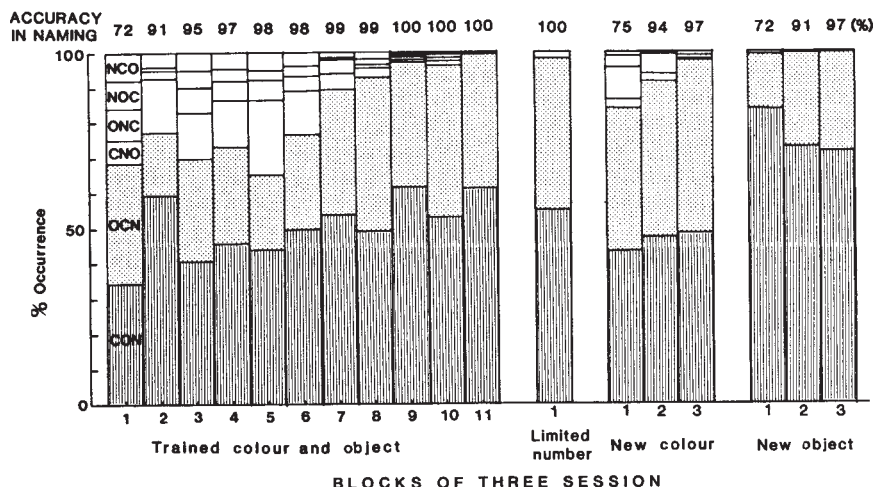


Fig. 3 Percentage of six possible sequences in three-term naming. N, number; C, colour; O, object. For example, NCO refers to trials in which the sequence 'number/colour/object' was used correctly. The open bars show the percentage of trials on which number was the first or second attribute named. The percentage of trials on which all three attributes of the sample items were named correctly is shown at the top of each column. The left panel shows the change of use of the six possible sequences as a function of three-term-naming training. A session ended when 100 rewards had been delivered. Each reward required two consecutive correct responses. Each block consists of three consecutive sessions. In the final block of three sessions of the task, the chimpanzee received 604 problems during 127 min and made only three mistakes. To investigate the determinant of the particular sequences, three kinds of tests were done. In the first test ('limited number'), the number of sample items was not varied but restricted to one of the five alternatives within a session. The objects and colours used in this first test were the same as those used in previous training. In the second test ('new colour'), new colours were introduced. The sample items consisted of the numbers from one to five, the five trained objects and five new colours, pink, brown, purple, white and grey. In the third test ('new object'), new objects were introduced. The sample items consisted of the five trained colours and five new objects, padlock, glove, glass, ball and key. In this test, the number of items to be named was restricted to one and two. Although those new samples had never been used in three-term naming, or even in simple numerical naming, the chimpanzee correctly described the three attributes of the samples. Decreasing the difficulty of numerical naming by using the same number of items in all samples in a given session, or increasing the difficulty of object or colour naming by introducing new colours and objects, had little effect on the favoured sequences. The chimpanzee seemed to have established the order of describing the three attributes of the sample items. However, the introduction of new samples had a temporary impact on the ratio of these two sequences. The introduction of new colours increased one sequence, object/colour/number. The introduction of new objects increased the use of the other sequence, colour/object/number. As those new samples were presented repeatedly, the ratio of the two favoured sequences gradually returned to the original level.



window and then pressed a start switch to light a set of numeric keys on a console. Correct and incorrect key choices were followed by different sounds. A piece of apple or a raisin was delivered after a number of consecutive correct responses. A session ended when 100 such rewards had been delivered or when 60 min had elapsed, whichever came first. Ai received one or two sessions a day. On the average, 303 trials were given in a 43-min session, with trials spaced ~8.5 s apart.

At the beginning of the first stage of training, only two numbered keys were available and only one or two red pencils were

shown. The number of red pencils and corresponding number keys was increased successively. Training was continued on each number set until accuracy exceeded 90% in at least two consecutive sessions. Acquisition of number naming with red pencils only as exemplars is shown in Fig. 2. The introduction of a new number was always accompanied by poorer identification of the maximum number learned in the previous stage. For example, in the course of initial training to name six items in addition to 1, 2, 3, 4 and 5, Ai made 274 errors out of 3,004 trials, 63.1% of these confusion between 5 and 6. Moreover,

Table 1 Generalization of numerical naming to a new colour, a new object or both a new colour and object in a chimpanzee

Numbers	Accuracy on probe trials			Accuracy by chance	Accuracy of trained samples on probe trials
	New colour, trained object	Trained colour, new object	New colour, new object		
1, 2	0.83 (5/6) Blue Pencil	0.50 (3/6) Red Paper	0.50 (3/6) Blue Paper	0.50	0.97
1, 2, 3	0.60 (18/30)* Yellow Pencil/paper	0.33 (10/30) Red/blue Brick	0.33 (5/15) Yellow Brick	0.33	0.99
1, 2, 3, 4	0.71 (34/48)‡ Black Pencil/paper/brick	0.54 (26/48)† Red/blue/yellow Spoon	0.56 (9/16)* Black Spoon	0.25	0.99
1, 2, 3, 4, 5	0.79 (79/100)‡ Green Pencil/paper/brick/spoon	0.57 (57/100)‡ Red/blue/yellow/black Tooth-brush	0.56 (14/25)* Green Tooth-brush	0.20	0.98

The generalization was tested by inserting probe trials in which the new colour, new object or both, was used to illustrate the number. These probe trials were included among trials during which trained samples were used. On trials with trained samples, the chimpanzee was informed whether her choice was correct or incorrect. On probe trials, which were inserted once every 10 trials on the average, the choice response was followed by neither the 'correct' nor 'error' feedback sound. Proportion correct is shown for the trained items as well as for the new colour and/or object introduced in each stage of training. The numbers in parentheses show the number of trials on which the proportions are based. The last column shows the proportion correct on the trials for the trained sample items, including all the colours and objects previously used. Generalization of the number name to new objects of familiar colour, to new colours of familiar objects, and to new objects in a new colour was at first poor. Less generalization was observed with new objects than new colours in each stage of training. As additional numbers were introduced and new objects and colours were used as exemplars, the correct number names were applied not only to new colours but also to new objects.

* $P < 0.05$; † $P < 0.01$; ‡ $P < 0.001$ significantly above the chance level (χ^2 test).

most errors (94.9%) were confusions between neighbouring numbers.

Following acquisition of a new number name, generalization to a new object and/or a new colour was tested. The specific objects and colours used as well as test performance are summarized in Table 1 and reported in more detail elsewhere¹³.

Accuracy during the final two sessions of naming 1, 2, 3, 4 and 5 with all five colours and five objects exceeded 98.5% in 830 trials. To reach this level of performance, a total of 95 sessions were required. In other words, 68 h 21 min (28,799 trials) elapsed from the beginning of training. It took an additional nine sessions (6 h 12 min, 3,004 trials) for Ai to reach criterion in selecting the correct Arabic numeral for groups of up to six red pencils.

How could Ai's newly acquired numerical skills be combined with object/colour naming? To answer this question, the chimpanzee was required to name not only the number of items, but also the colour and objects used to illustrate the number.

Fifteen keys, each of which represented one of five numbers (1, 2, 3, 4, 5), five colours (red, blue, yellow, green, black) and five objects (pencil, paper, brick, spoon, tooth-brush), were operative at one time. The sequence in which the keys were depressed was free, but the chimpanzee was required to select the three keys correctly describing each of the three attributes of $5 \times 5 \times 5 = 125$ types of sample items. For example, when five blue tooth-brushes were shown in the display window as a sample item, it was necessary for Ai to press keys of '5', 'blue' and 'tooth-brush' in any order. When the chimpanzee pressed an incorrect key, the 'error' sound followed and the trial ended. The pressed key was darkened and inoperative, so that the repeated response such as 'red/red' was inhibited.

Although no particular 'word order' was required, the chimpanzee favoured two particular sequences among the six possible alternatives, colour/object/number and object/colour/number (left panel of Fig. 3). In both sequences, numerical naming was always last. In the final sessions, it generally took less than 3 s for Ai to describe the object, colour and number of the sample items. She seldom looked back at the display between each component response. The response latency in the first component was the longest and the third component (numerical naming in almost all cases) was longer than the second one. The previously favoured sequences continued to be used for new samples with new colours and/or new objects (details in the right panel of Fig. 3).

These results suggest that the readiness with which the attributes are named is a partial determinant of the favoured sequences. Accuracy in numerical naming was always lower than that in object or colour naming in the three-term naming task. Moreover, number names were learned more slowly than were the names of objects and colours. It took 1,052 trials (171 min) for Ai to reach criterion in her first object discrimination task ('padlock' and 'glove'), 538 trials (105 min) for the first colour discrimination ('red' and 'green') and 1,821 trials (253 min) for the first number discrimination ('1' and '2'). In the previous stage of two-term naming, Ai had experienced both sequences, object/colour and colour/object. The previous experience may be the factor facilitating the reversible bonding of object and colour names in three-term naming. It seems that the favoured sequences in three-term naming are determined by complex factors, which include readiness with which the attributes are named and her previous experiences.

The range of numbers (1-6), the variety of sample items (at least 125 types), generalization to new items and the complex sequence of responses make it highly unlikely that the naming depends on the mere association of particular sample items and alternative keys. The present study suggests that the chimpanzee was able to match the sample items on the basis of numerosity in addition to colour and types of object.

I thank Drs Kiyoko Murofushi and Toshio Asano for helpful discussions and support; Drs Pamela Asquith and Sheila Chase for suggesting improvements to the original manuscript. This

work was supported by a grant-in-aid from the Ministry of Education, Science and Culture, Japan (240008, 58400006, 58710059).

Received 30 October 1984; accepted 22 January 1985.

1. Gardner, R. A. & Gardner, B. T. *Science* **165**, 664-672 (1969).
2. Premack, D. *Intelligence in Ape and Man* (Erlbaum, Hillsdale, New Jersey, 1976).
3. Rumbaugh, D. M. (ed.) *Language Learning by a Chimpanzee* (Academic, New York, 1977).
4. Fouts, R. S. *Science* **180**, 978-980 (1973).
5. Patterson, F. G. *Brain and Language* **5**, 72-97 (1978).
6. Terrace, H. S., Petitto, L. A., Sanders, R. J. & Bever, T. G. *Science* **206**, 891-902 (1979).
7. Ferster, C. B. & Hammer, C. E. Jr in *Operant Behavior: Areas of Research and Application* (ed. Honig, W. K.) 634-676 (Appleton-Century-Crofts, New York, 1966).
8. Dooley, G. B. & Gill, T. V. in *Language Learning by a Chimpanzee* (ed. Rumbaugh, D. M.) 247-260 (Academic, New York, 1977).
9. Woodruff, G. & Premack, D. *Nature* **293**, 568-570 (1981).
10. Davis, H. & Memmott, J. *Psychol. Bull.* **92**, 547-571 (1982).
11. Asano, T., Kojima, T., Matsuzawa, T., Kubota, K. & Murofushi, K. *Proc. Japan. Acad.* **58B**, 118-122 (1982).
12. Kojima, T. *Int. J. Primat.* **5**, 161-182 (1984).
13. Matsuzawa, T., Asano, T., Kubota, K. & Murofushi, K. in *Current Perspectives in Primate Social Dynamics: Selected Proc. 9th Congr. int. primatol. Soc.* (eds Taub, D. M. & King, F. A.) (Van Nostrand Reinhold, New York, in the press).
14. Matsuzawa, T. *J. hum. Evol.* (in the press).

Co-localization of corticotropin releasing factor and vasopressin mRNA in neurones after adrenalectomy

Betty Wolfson, Robert W. Manning, Leonard G. Davis, Rene Arentzen & Frank Baldino Jr*

Central Research & Development Department, Experimental Station, E. I. du Pont de Nemours & Co., Wilmington, Delaware 19898, USA

The discrete anatomical distribution of arginine vasopressin and corticotropin releasing factor (CRF) immunoreactivity in the paraventricular nucleus (PVN) of the rat hypothalamus is altered after adrenalectomy¹⁻³. Not only is the immunostaining of both peptides enhanced, but vasopressin immunoreactivity, normally confined to the magnocellular subdivision, becomes clear in a large percentage of CRF neurones in the parvocellular subdivision². These changes in immunoreactivity may reflect changes in post-translational events, peptide metabolism or genomic activity that lead indirectly or directly to the enhanced expression of vasopressin. Here we report that levels of transcripts homologous to vasopressin messenger RNA increase in the PVN after adrenalectomy, in parallel with increases in vasopressin immunoreactivity. In fact, after adrenalectomy, vasopressin mRNA can be detected in CRF-immunoreactive neurones. These results indicate that a considerable degree of plasticity is retained by the adult neuronal genome of the rat and that this plasticity may be modulated by the endocrine environment.

CRF and vasopressin act synergistically to release adrenocorticotrophic hormone (ACTH) from the anterior pituitary⁴. Recently, immunohistochemical studies have demonstrated that CRF-containing fibres in the median eminence arise from cells in the parvocellular subdivision of the PVN⁵, whereas the greatest density of vasopressin-containing neurones is found in the magnocellular subdivision of the PVN⁶. After adrenalectomy, the anatomical distribution of these peptides becomes less discrete; vasopressin and CRF can be found co-localized in >70% of parvocellular neurones². It has been suggested that the removal of adrenal steroids accounts for these changes in neuropeptide expression⁷. Because steroids regulate peptide synthesis by modulating genomic activity⁸, it seemed possible that adrenal hormones regulate neuronal synthesis of vasopressin by influencing some process preceding mRNA translation. Therefore, we examined the effects of adrenalectomy on vasopressin mRNA levels in the PVN by applying *in situ* hybridization techniques in conjunction with standard

* To whom correspondence should be addressed.

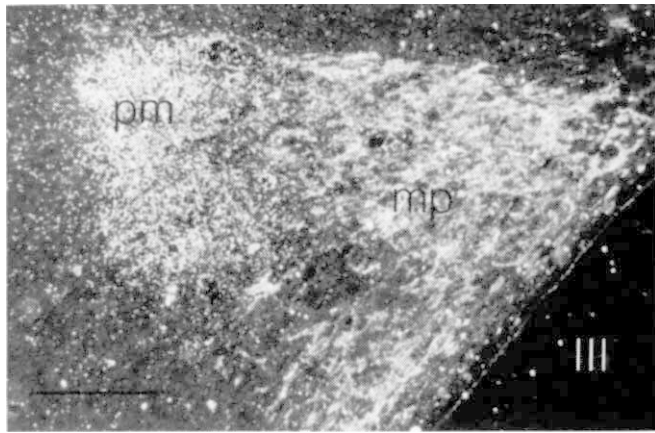


Fig. 1 Simultaneous detection of vasopressin mRNA and CRF-like immunoreactivity in the PVN of a control rat photographed under dark-field illumination. CRF-like immunoreactive neurones are confined to the medial parvocellular division (mp), whereas vasopressin mRNA (silver grains) is largely confined to the posterior magnocellular division (pm). III, Third ventricle. Scale bar, 200 μ m.

Methods. Colchicine-treated (75 μ g) rats were perfused with phosphate buffered saline (pH 7.0) followed by buffered 10% formalin containing 15% sucrose. Transverse 35- μ m sections containing the PVN were cut on a freezing microtome and processed for CRF immunoreactivity using the avidin-biotin method⁹ and diaminobenzidine as the chromogen. Immunostained sections were then delipidated in ethanol, rehydrated, rinsed in diethylpyrocarbonate (1 μ l 5 ml⁻¹), slide-mounted and incubated overnight at room temperature with ³²P-labelled probe (5×10^4 c.p.m. per section) diluted in hybridization buffer (2 $\times$ SSC, 10 $\times$ Denhardt's, 50% formamide, 0.1% SDS and 0.1% denatured salmon sperm DNA). The next day, slides were rinsed in 2 $\times$ SSC for 4 h with solution changes every 15 min. Slides were then air dried and dipped in Ilford L4 nuclear research emulsion diluted 1:1 with water. Because this emulsion is fine grained, slides were exposed for 4 weeks in light-proof boxes at 4°C. Development was performed with Kodak D-19 (diluted 1:1) for 5 min at 20°C and followed by fixation in 30% sodium thiosulphate.

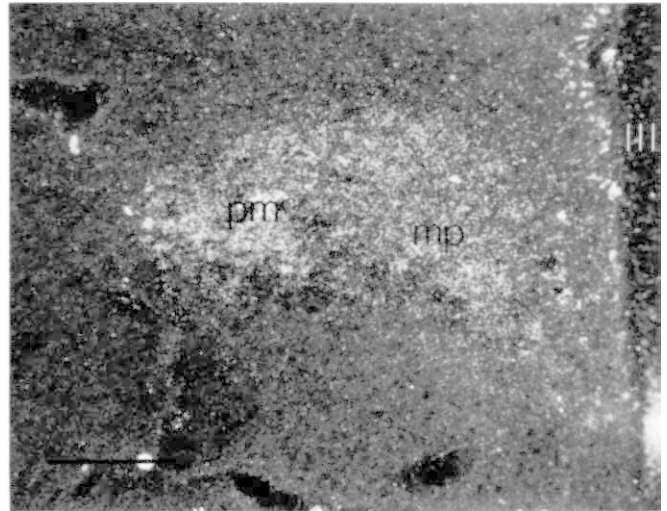


Fig. 2 Localization of vasopressin mRNA in the parvocellular PVN of an adrenalectomized rat photographed under dark-field illumination. The highest density of grains is still seen in the posterior magnocellular division (pm), but there is also a considerable amount of hybridized probe in the medial parvocellular division (mp). III, Third ventricle. Scale bar, 230 μ m.

Methods. A 48-base oligonucleotide complementary to the glycoprotein region of vasopressin mRNA was prepared according to previously published methods¹⁰. Two different 24-mers were synthesized on an Applied Biosystems (Model 380A) DNA synthesizer and purified by reverse-phase HPLC on a μ Bondapak C18 column (Waters Associates). One of the 24-mers was labelled at its 5' end with ³²P by reaction with [γ -³²P]ATP and T4 polynucleotide kinase. The 3' end of this 24-mer was then ligated to the 5' end of the second 24-mer, phosphorylated with ATP and T4 polynucleotide kinase, with the aid of a 17-mer holding fragment and T4 DNA ligase. The resulting 48-mer was separated from excess reactants and the complementary holding fragment by polyacrylamide gel electrophoresis in denaturing conditions. The final activity of the probe was $\sim 10^8$ c.p.m. μ g⁻¹.

immunocytochemical methods. (For simplicity, the terms vasopressin mRNA and CRF- or vasopressin-immunoreactivity used here imply their putative nature.)

Adrenalectomized (5 days) and sham-operated control male Sprague-Dawley rats were colchicine-treated, perfused 48 h later and processed for vasopressin or CRF immunostaining on alternate sections, using the avidin-biotin method⁹ with commercially produced antisera (Immunonuclear). *In situ* hybridization techniques with radiolabelled, synthetic oligodeoxyribonucleotide probes were then applied to detect vasopressin mRNA in the immunostained sections. The hybridization procedures used and controls for the specificity of the resultant signal were, with minor modifications, as described previously¹⁰. Conventional autoradiographic procedures were used to visualize hybridized probe. No differences in the hybridization signal were observed between colchicine- and non-colchicine-treated animals.

In sections from control animals, vasopressin immunoreactivity and the hybridization signal were largely confined to the magnocellular subdivision of the PVN, as expected if the probe specifically hybridized to transcripts homologous to vasopressin mRNA (Fig. 1). This localization of vasopressin mRNA was seen in all control tissue, regardless of whether vasopressin or CRF was immunostained in the section. Since vasopressin and CRF are normally distributed discretely in the PVN, the consistency of the distribution of the hybridization signal indicates that the immunoreaction product did not lead to the production of false-positive chemographic images.

Adrenalectomy induced the appearance of vasopressin immunoreactivity in dorsal and medial regions of the parvocellular PVN. Hybridized probe could also be detected in these

same parvocellular regions (Fig. 2). Indeed, vasopressin mRNA could be seen in the cytoplasm of parvocellular CRF-immunoreactive neurones (Fig. 3). Adrenalectomy did not lead to the appearance of vasopressin mRNA in CRF neurones in all regions of the central nervous system (CNS) (data not shown).

These results demonstrate the ability of the endocrine environment to regulate transcript levels in a discrete region of the adult CNS. Certainly, the changes in genomic expression described here need not have been a direct consequence of decreased circulating adrenal hormones. The loss of feedback inhibition of both pituitary corticotrophs¹¹ and CRF-secreting neurones⁵ upon adrenalectomy leaves open the possibility that elevated concentrations of ACTH or CRF contribute to the enhanced production of vasopressin mRNA. CRF has been shown to contribute significantly to the elevation of proopiomelanocortin mRNA levels seen in the pars intermedia of adrenalectomized rats¹². Furthermore, changes in neuronal activity that might occur in response to endocrine manipulations^{13,14} could themselves lead to changes in neuropeptide expression. For example, chemical depletion of adrenergic terminals in the hypothalamus leads to enhanced CRF immunoreactivity in the PVN¹⁵; depolarization of sympathetic ganglia is known to result in changes in their neuropeptide content¹⁶. Whether such indirect actions consequent to adrenalectomy also affect the observed redistribution of vasopressin mRNA remains to be determined. In contrast to the increase in vasopressin immunoreactivity reported here in the PVN after adrenalectomy, a dramatic decrease in vasopressin immunoreactivity in the adult rat habenula has been found after gonadectomy¹⁷. Whether this effect of gonadectomy reflects changes in genomic activity

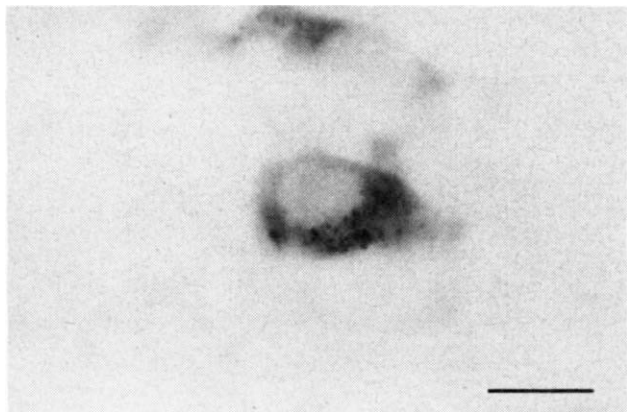


Fig. 3 CRF-immunoreactive neurone containing vasopressin mRNA. Note the confinement of the silver grains to the cytoplasm. Scale bar, 20 μ m.

remains to be demonstrated. In any case, the observation that regional levels of the same neuropeptide may be differentially regulated by different humoral factors is particularly striking.

Our results show that changes in the endocrine composition caused by adrenalectomy promote the expression of vasopressin mRNA in parvocellular regions of the rat PVN. Of particular note is the induction of vasopressin mRNA in CRF-immunoreactive neurones under these conditions. These results demonstrate that differential regulation of gene expression occurs in the adult CNS in response to physiological stimuli.

We thank T. O'Kane for technical assistance, Dr S. A. Tomiko for critical comments and Y. J. King for typing the manuscript.

Received 19 November 1984; accepted 8 February 1985.

1. Kiss, J. Z., Mezey, E. & Skirboll, L. *Proc. natn. Acad. Sci. U.S.A.* **81**, 1854-1858 (1984).
2. Sawchenko, P. E., Swanson, L. W. & Vale, W. W. *Proc. natn. Acad. Sci. U.S.A.* **81**, 1883-1887 (1984).
3. Tramu, G., Croix, C. & Pillez, A. *Neuroendocrinology* **37**, 467-469 (1983).
4. Rivier, C. & Vale, W. *Endocrinology* **113**, 939-942 (1983).
5. Swanson, L. W., Sawchenko, P. E., Rivier, J. & Vale, W. W. *Neuroendocrinology* **36**, 165-186 (1983).
6. Sawchenko, P. E. & Swanson, L. W. *J. comp. Neurol.* **205**, 260-272 (1982).
7. Sawchenko, P. E. & Swanson, L. W. *Neurosci. Abstr.* **10**, 83 (1984).
8. McEwen, B. S., Gerlach, J. L., Luine, V. N. & Lieberburg, I. *Psychoneuroendocrinology* **2**, 249-255 (1977).
9. Hsu, S.-M., Raine, L. & Fanger, H. *J. Histochem. Cytochem.* **29**, 577-580 (1981).
10. Arentzen, R. *et al.* *J. cell. Biochem.* (in the press).
11. Haynes, R. C. Jr & Murad, F. in *The Pharmacological Basis of Therapeutics* (eds Goodman Gilman, A., Goodman, L. S. & Gilman, A.) 1469 (Macmillan, New York, 1980).
12. Bruhn, T. O., Sutton, R. E., Rivier, C. L. & Vale, W. W. *Neuroendocrinology* **39**, 170-175 (1984).
13. Mandelbrod, I., Feldman, S. & Werman, R. *Brain Res.* **80**, 303-315 (1974).
14. Papir-Kricheli, D. & Feldman, S. *Neuroscience* **7**, 2793-2800 (1982).
15. Mezey, E., Kiss, J. Z., Skirboll, L. R., Goldstein, M. & Axelrod, J. *Nature* **310**, 140-141 (1984).
16. Adler, J. E., Kessler, J. A. & Black, I. B. *Dev. Biol.* **102**, 417-425 (1984).
17. DeVries, G. J., Buijs, R. M. & Sluiter, A. A. *Brain Res.* **298**, 141-145 (1984).

Stimulation of connective tissue cell growth by substance P and substance K

Jan Nilsson*, Anne M. von Euler†
& Carl-Johan Dalsgaard‡

Departments of *Histology and ‡Anatomy, Karolinska Institutet, and †Wenner-Gren Institute, University of Stockholm, Stockholm, Sweden

Connective tissue cells proliferate actively when cultured in the presence of serum. Platelet-derived growth factor (PDGF), a basic protein of relative molecular mass ~30,000, has been identified as the major serum mitogen for these cells¹; its main physiological/pathophysiological role may be to initiate wound healing in connection with tissue injury. However, growth of cultured cells

is also influenced by several other factors, including epidermal growth factor², fibroblast growth factor³, insulin and somatomedins⁴. Furthermore, Rozengurt and Sinnett-Smith recently showed that bombesin, a neuroendocrine peptide isolated from frog skin, stimulates DNA synthesis and cell division in cultures of a specific subtype of 3T3 cells⁵. Substance P and substance K (also known as neurokinin A or neuromedin L) are mammalian peptides belonging to the tachykinin family⁶. Substance P has been studied extensively; it is distributed widely throughout the central and peripheral nervous system⁷, including primary sensory neurones⁸⁻¹⁰, and can be released in the periphery from axon collaterals of stimulated pain fibres^{11,12} and contribute to the inflammatory response¹³. Substance K is a member of the tachykinin family isolated from mammalian spinal cord^{14,15}; Nawa *et al.*¹⁶ determined the primary structure of two types of substance P precursors, one of which contained a sequence homologous to substance K, as well as the sequence of substance P. We report here that substance P and substance K stimulate DNA synthesis in cultured arterial smooth muscle cells and human skin fibroblasts, and that this stimulation is inhibited by the substance P-antagonist spantide¹⁷.

Smooth muscle cells were isolated enzymatically from the aortic media of adult Sprague-Dawley rats¹⁸ and fibroblasts were isolated by outgrowth from human skin explants¹⁹. Secondary cultures of muscle cells and cultures of fibroblasts in the 10th to 12th passage were growth-arrested by transfer to serum-free medium for 24 h. The rate of entrance of the cells into S phase of the cell cycle was determined by counting the number of autoradiographically labelled nuclei after a 24-h incubation with ³H-thymidine. Exposure to medium containing 10% newborn calf serum initiated DNA synthesis in $71.8 \pm 3.8\%$ ($\pm$ s.d.) of the muscle cells and in $47 \pm 5.3\%$ of the fibroblasts, compared with $18.6 \pm 3.0\%$ and $17.1 \pm 2.7\%$, respectively, in unstimulated cultures.

Addition of substance P or substance K to serum-free medium stimulated DNA synthesis in cultured muscle cells in a dose-dependent manner, whereas addition of bombesin had no effect (Figs 1a, 2). Substance K was a more potent stimulator than substance P, at 10^{-6} M producing $62.0 \pm 1.4\%$ labelled nuclei while 10^{-5} M substance P produced $38.3 \pm 4.4\%$ labelled nuclei. PDGF at 25 ng ml^{-1} produced maximal stimulation of DNA synthesis in muscle cells grown in serum-free medium ($42.6 \pm 4.7\%$ labelled nuclei; see ref. 18), whereas 2 ng ml^{-1} PDGF produced only a small increase (Fig. 1a). DNA synthesis in muscle cells stimulated with 2 ng ml^{-1} PDGF could be further stimulated by the addition of substance P or K. However, the combined effect of substance P or K and suboptimal concentrations of PDGF was less than the sum of their individual effects. Furthermore, at concentrations $>10^{-8}$ M, substance K was more active alone than in combination with suboptimal concentrations of PDGF (Fig. 1a).

Substance P and substance K also stimulated initiation of DNA synthesis in fibroblast cultures; no effect was observed after treatment with bombesin (Fig. 1b). Again, substance K was more potent than substance P. The rate of DNA synthesis in fibroblasts stimulated with suboptimal concentrations of PDGF was further enhanced after treatment with substance P or substance K, but the enhancement was lower than the individual effects of either substance (Fig. 1b).

To determine the specificity of the stimulating action of substance P and substance K, we performed control experiments with cultures preincubated with spantide ([D-Arg¹, D-Trp^{7,9}, Leu¹¹]-substance P; Bachem), a substance P antagonist¹⁷. We found that the addition of spantide (3×10^{-6} M- 3×10^{-5} M) to serum-free medium caused a limited stimulation of DNA synthesis, 3×10^{-5} M producing $26.4 \pm 2.0\%$ labelled nuclei, and a significant reduction of DNA synthesis induced by 10^{-7} M of substance P or substance K in both muscle cells and human fibroblasts. Spantide did not reduce the rate of DNA synthesis in cultures stimulated with 25 ng ml^{-1} PDGF (Table 1).

There is accumulating evidence that many factors are involved in the regulation of cell growth. The present results demonstrate

Table 1 Effect of the substance P analogue spantide on DNA synthesis induced by substance P (SP), substance K (SK) and PDGF

	Smooth muscle cells		Human fibroblasts	
	% Labelled nuclei	s.d.	% Labelled nuclei	s.d.
MCDB 104	18.6	3.0	17.1	2.7
MCDB 104/ 3×10^{-6} M spantide	17.2	2.7	15.7	3.2
MCDB 104/ 10^{-5} M spantide	21.1	2.1	23.8	0.4
MCDB 104/ 3×10^{-5} M spantide	26.4	2.0	26.1	2.3
10^{-7} M SP	40.9	0.6	32.0	1.1
10^{-7} M SP/ 3×10^{-6} M spantide	32.0*	2.2	28.7	2.0
10^{-7} M SP/ 10^{-5} M spantide	33.9	5.3	28.5*	0.4
10^{-7} M SP/ 3×10^{-5} M spantide	25.8	1.2	24.6	0.2
10^{-7} M SK	50.1	1.3	38.2	1.4
10^{-7} M SK/ 3×10^{-6} M spantide	41.5*	0.8	28.7*	3.4
10^{-7} M SK/ 10^{-5} M spantide	33.3	4.2	27.2	2.9
10^{-7} M SK/ 3×10^{-5} M spantide	27.6	1.6	23.1	1.2
PDGF 25 ng ml $^{-1}$	40.8	3.3	45.9	4.3
PDGF 25 ng ml $^{-1}$ / 3×10^{-6} M spantide	44.5	4.1	45.0	5.0
PDGF 25 ng ml $^{-1}$ / 10^{-5} M spantide	41.1	1.1	43.2	1.3
PDGF 25 ng ml $^{-1}$ / 3×10^{-5} M spantide	41.9	0.9	43.2	2.5

Medium MCDB 104 (Gibco) was used as control. The fraction of labelled nuclei was determined as described in Fig. 1 legend. Each value represents the mean of triplicate cultures.

* The first value in each group, significantly different ($P < 0.05$) from the control.

that substance P and substance K, two structurally related neurotransmitters found in both the central and peripheral nervous system, may act as mitogens for connective tissue cells. The addition of 10^{-7} M of substance P alone or together with suboptimal concentrations of PDGF stimulated initiation of DNA synthesis, whereas substance K stimulated entrance to S phase at concentrations as low as 10^{-9} M. Note that substance K was 100 times more potent in stimulating DNA synthesis than substance P, indicating an action via a substance P-E receptor subtype^{20,21}. Substances P and K were less effective in stimulating DNA synthesis if the cells were simultaneously stimulated with suboptimal concentrations of PDGF. This lack of potentiating effect may result from the existence of similarities in the mechanisms of action of neuropeptides and growth factors. Indeed, our results indicate that PDGF decreases the stimulatory effect of substance K in smooth muscle cells, which suggests a complex interaction between the mitogenic effect of PDGF and substance K. DNA synthesis induced by substance P or substance K was markedly reduced by incubation with spantide, an analogue of substance P with antagonistic activity, suggesting a competitive inhibition of receptor binding. Spantide alone had a small stimulatory effect, consistent with the previously reported agonistic activity of the substance P analogues⁷. It is interesting that bombesin, which has two C-terminal amino acids in common with substance P and substance K, failed to induce DNA synthesis. We therefore conclude that the stimulatory effect is mediated via specific tachykinin receptors.

The mechanisms by which substance P and substance K stimulate DNA synthesis remain to be determined. However, it

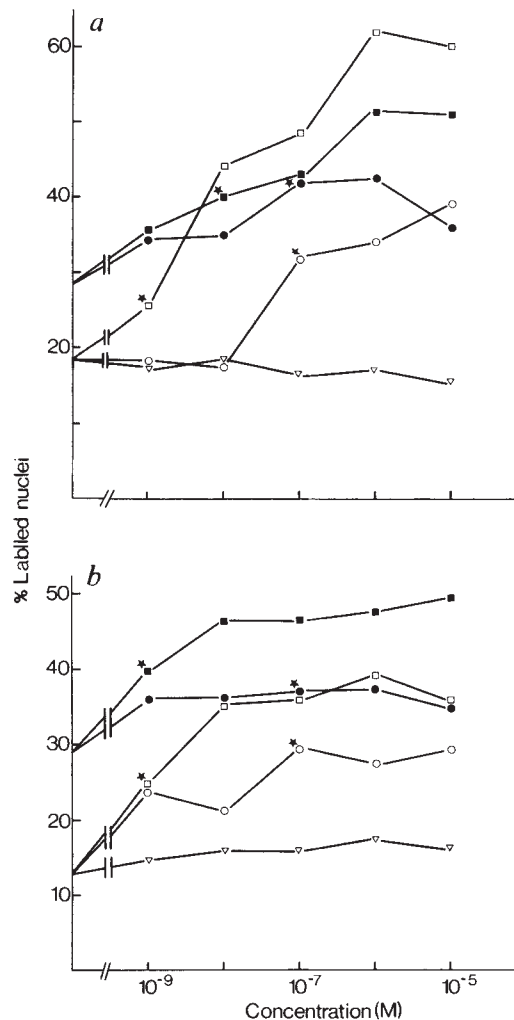


Fig. 1 Effects of substance P and substance K on initiation of DNA synthesis in smooth muscle cells (a) and skin fibroblasts (b). Subconfluent, secondary cultures of muscle cells were growth-arrested by transfer to medium MCDB 104/0.1% bovine serum albumin (serum-free medium) for 24 h. They were then incubated with $2.5 \mu\text{Ci ml}^{-1}$ of ^3H -thymidine and various concentrations of substance P (Bachem) alone ($\circ$) or with ($\bullet$) 2 ng ml^{-1} of PDGF²⁵ added, various concentrations of substance K (Peninsula) alone ($\square$) or with ($\blacksquare$) 2 ng ml^{-1} of PDGF or with various concentrations of bombesin (Peninsula) alone (∇) added for another 24 h. The cells were fixed in 3% buffered glutaraldehyde and processed for autoradiography as described elsewhere¹⁸. The fraction of labelled nuclei was determined by counting 500 randomly selected cells in each specimen. Values are given as means of triplicate cultures (s.d. $< 20\%$). * Indicates the first value in each group, which is significantly different ($P < 0.05$) from the control.

is interesting that both substance P²² and PDGF²³ have been found to stimulate hydrolysis of phosphatidylinositol 4,5-bisphosphate into diacylglycerol and inositol trisphosphate and that inositol trisphosphate has recently been shown to function as a second messenger for mobilization of intracellular calcium²⁴, which is believed to participate in regulation of cell proliferation.

The present data suggest a possible role for tachykinins in wound healing. It is tempting to speculate that substance P and/or substance K released from peripheral neurones in connection with tissue injury not only contribute to the inflammatory response but also could help other mitogens, such as PDGF, to stimulate proliferation of surrounding connective tissue cells and thus initiate a healing response.

Fibroblasts were kindly supplied by Dr Ove Ceder and Dr Hans Kollberg; PDGF was a gift from Dr Carl-Henrik Heldin. This work was supported by the Swedish MRC (06537 and

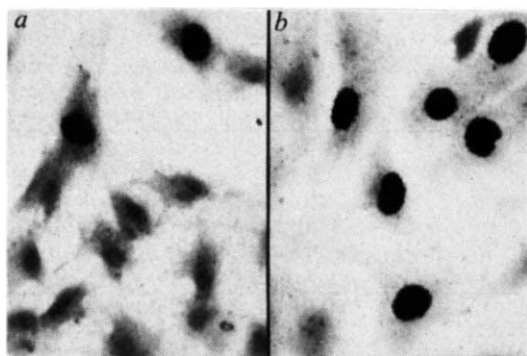


Fig. 2 Micrographs of representative fields of autoradiographs from a culture of smooth muscle cells treated with serum-free medium (a) or 10^{-6} M substance K (b). $\times 600$.

07126), the King Gustaf V 80th Birthday Fund, the funds of Karolinska Institutet, Magnus Bergvalls Stiftelse, Tore Nilssons Fond, Harald & Greta Jeansson's Stiftelse, the Swedish National Association against Heart and Lung Disease and SRA.

Received 19 October 1984; accepted 11 February 1985.

1. Westermark, B. *et al.* in *Growth and Maturation Factors* (ed. Guroff, G.) 73–116 (Wiley, New York, 1983).
2. Carpenter, G. & Cohen, S. A. *Rev. Biochem.* **48**, 193–216 (1979).
3. Gospodarowicz, D., Greenburg, G., Bialecki, H. & Zetter, B. R. *In Vitro* **14**, 85–118 (1978).
4. Van Wyk, J. J., Furianetto, R. W., Plet, A. S., D'erciole, A. J. & Underwood, L. E. *Natn. Cancer Inst. Monogr.* **43**, 141–147 (1978).
5. Rozengurt, E. & Sinnet-Smith, J. *Proc. natn. Acad. Sci. U.S.A.* **80**, 2936–2940 (1983).
6. Erspamer, V. *Trends Neurosci.* **11**, 267–269 (1981).
7. Pernow, B. *Pharmac. Rev.* **35**, 85–141 (1983).
8. Hökfelt, T., Kellert, J.-O., Nilsson, G. & Pernow, B. *Science* **190**, 889–890 (1975).
9. Cuello, A. C., Del Fiacco, M. & Paxinos, G. *Brain Res.* **152**, 499–509 (1978).
10. Dalsgaard, C.-J., Jonsson, C.-E., Hökfelt, T. & Cuello, A. C. *Experientia* **39**, 1018–1020 (1983).
11. Brodin, E., Gazelius, B., Olgart, L. & Nilsson, G. *Acta physiol. scand.* **111**, 141–149 (1981).
12. Bill, E., Stjernschantz, J., Maudahl, A., Brodin, E. & Nilsson, G. *Acta physiol. scand.* **106**, 371–373 (1979).
13. Lembeck, F. & Holzer, P. *Naunyn-Schmiedeberg's Archs Pharmacol.* **310**, 175–183 (1979).
14. Kimura, S., Okada, M., Sugita, Y., Kanazawa, I. & Munekata, E. *Proc. Japan. Acad.* **59**, 101–104 (1983).
15. Minamino, N., Kangawa, K., Fukuda, A. & Matsuo, H. *Neuropeptides* **4**, 157–164 (1984).
16. Nawa, H., Hirose, T., Takashima, H., Inayama, S. & Nakanishi, S. *Nature* **306**, 32–36 (1984).
17. Rosell, S., Björkeröth, U., Xu, J.-C. & Folkers, K. *Acta physiol. scand.* **117**, 445–449 (1983).
18. Nilsson, J., Ksiazek, T., Heldin, C.-H. & Thyberg, J. *Expl Cell Res.* **145**, 231–237 (1983).
19. Roomans, G. M., Ceder, O. & Kollberg, H. *Ultrastruct. Path.* **2**, 53–58 (1981).
20. Lee, C.-M., Iversen, L. L., Hanely, M. R. & Snadberg, B. E. *Naunyn-Schmiedeberg's Archs Pharmacol.* **318**, 281–287 (1982).
21. Hunter, J. C. & Maggio, J. E. *Eur. J. Pharmacol.* **97**, 150–160 (1984).
22. Berridge, M. J., Dawson, R. M. C., Downes, C. P., Heslop, J. P. & Irvine, R. F. *Biochem. J.* **212**, 437–482 (1983).
23. Berridge, M. J., Heslop, J. P., Irvine, R. F. & Brown, K. D. *Biochem. J.* **222**, 195–201 (1984).
24. Berridge, M. J. & Irvine, R. F. *Nature* **312**, 315–321 (1984).
25. Heldin, C.-H., Westermark, B. & Wasteson, A. *Biochem. J.* **193**, 907–913 (1981).

Voltage dependence of Na/K pump current in isolated heart cells

David C. Gadsby*, Junko Kimura & Akinori Noma

National Institute for Physiological Sciences, Okazaki, 444 Japan
*Laboratory of Cardiac Physiology,
The Rockefeller University, New York, New York 10021, USA

The Na/K pump usually pumps more Na^+ out of the cell than K^+ in, and so generates an outward component of membrane current^{1,2} which, in the heart, can be an important modulator of the frequency and shape of the cardiac impulse^{3–5}. Because it is electrogenic, Na/K pump activity ought to be sensitive to membrane potential, and it should decline with hyperpolarization^{6–9}. However, such voltage dependence of outward pump current has yet to be demonstrated^{6–10}, one reason being the technical difficulty of accurately measuring pump current over a sufficiently wide voltage range. The whole-cell patch-clamp technique¹¹ allows effective

control of both intracellular and extracellular solutions as well as membrane voltage. Applying this technique to myocardial cells isolated from guinea pig ventricle, we have measured Na/K pump current between -140 mV and $+60$ mV, after minimizing passive currents flowing through Ca^{2+} , K^+ and Na^+ channels. We report here that strongly activated pump current shows marked voltage dependence; it declines steadily from a maximal level near 0 mV, becoming very small at -140 mV. Pump current–voltage relationships will provide essential information for testing models of the Na/K pump mechanism and for predicting pump-mediated changes in the electrical activity of excitable cells.

We used wide-tipped patch pipettes to modify intracellular ionic composition so as to reduce outward K^+ currents¹², by lowering the pipette K^+ concentration ($[\text{K}^+]_{\text{pipette}}$) to 10 mM (K^+ replaced by Cs^+) and by adding 20 mM tetraethylammonium (TEA). To diminish inward currents in Ca^{2+} and K^+ channels, external solutions were nominally Ca^{2+} -free and included small amounts of Cd^{2+} , Ba^{2+} and Cs^+ . The resulting low membrane conductance facilitated the control of membrane potential over an extensive range and enhanced the signal-to-noise ratio for detection of Na/K pump current.

Rod-shaped, quiescent ventricular cells were obtained by collagenase digestion of guinea pig hearts, followed by enzyme-free incubation in high $[\text{K}^+]$, low $[\text{Ca}^{2+}]$ medium^{13,14}. The cells were superfused at 36°C with normal Tyrode's solution containing (in mM): 144 NaCl, 5.4 KCl, 0.3 NaH_2PO_4 , 1.8 CaCl_2 , 0.5 MgCl_2 , 5.5 dextrose, 5 HEPES (pH 7.4). Giga-ohm seals were obtained using wide-tipped (4 – 5 μm), fire-polished pipettes filled with the same Tyrode's solution (pipette resistance ~ 1 M Ω), then, before rupturing the membrane patch, the pipette solution was exchanged¹⁵ for one containing (in mM): 100 Cs-aspartate, 10 Tris₂ATP, 5 K₂-creatine phosphate, 1 NaH_2PO_4 , 20 TEA-Cl, 5.5 MgCl_2 , 5 EGTA, 5 HEPES (pH 7.4). After rupture of the patch, the holding potential was set at -40 mV to inactivate Na^+ channels and, after allowing ~ 5 min for the cell interior to equilibrate with the pipette solution (see Fig. 1A), the cells were superfused with nominally Ca^{2+} -free Tyrode's solution (CaCl_2 replaced by 1.8 mM MgCl_2 , 0.9 mM BaCl_2 , 0.2 mM CdCl_2 and 1 mM CsCl). In these conditions, membrane currents elicited by voltage steps were small and largely time-independent (Figs 1B, 2A), confirming that channel currents were depressed.

Figure 1A shows that a sudden increase in $[\text{Na}^+]_{\text{pipette}}$ from 1 mM to 34 mM (Na^+ replacing Cs^+) caused a rapid outward shift of holding current which was completely abolished by external application of 10 μM ouabain. The fact that ouabain specifically inhibits the Na/K pump suggests that, at -40 mV, the entire outward current shift caused by raising the pipette (and, hence, intracellular) $[\text{Na}^+]$ results from Na/K pump activation. To determine effects at other voltages, we applied 300 -ms clamp pulses to potential levels between -140 mV and $+60$ mV. Figure 1B shows samples of the currents obtained in the presence of high pipette $[\text{Na}^+]$ immediately before (a) and just after (b) pump inhibition by ouabain; c shows the ouabain-sensitive components obtained by digital subtraction of the corresponding traces in a and b. These difference traces already indicate that the size of the ouabain-sensitive current varies with membrane potential.

The detailed voltage dependence of the ouabain-sensitive current is illustrated in Fig. 2. The different symbols in Fig. 2A show current levels measured during the four data-collection periods numbered in Fig. 1A. The currents in the presence of ouabain ($\bullet$) are practically superimposable on those recorded with $[\text{Na}^+]_{\text{pipette}}$ of 1 mM (Δ), arguing that, over the entire voltage range examined, the outward current elicited by raising the intracellular $[\text{Na}^+]$ results from activation of the Na/K pump. The data points in Fig. 2B show the amplitudes of the difference currents, obtained by digital subtraction of the currents in Fig. 2A: triangles show the Na^+ -activated current, and the open and solid circles show the ouabain-sensitive current determined, respectively, by washing on, and washing off, ouabain. All three sets of points demonstrate that the ouabain-

Fig. 1 Current changes due to activation or inhibition of the Na/K pump in a ventricular cell superfused with 5.4 mM K^+ , Ca^{2+} -free Tyrode's solution. **A**, Chart recording: upper trace, membrane potential; middle trace, membrane current; the lower line indicates approximate timing of step changes in $[Na^+]_{\text{pipette}}$ between 1 and 34 mM, and the bar beneath it marks the exposure to 10 μM ouabain; the numbers above the voltage trace indicate acquisition of four sets of current-voltage data. To change the pipette solution¹⁵, negative pressure (~ -30 cm H_2O) draws fluid from a miniature reservoir through fine silastic tubing and delivers it to within 200 μm of the pipette tip via a narrow, tapered polythene catheter attached to a 25-gauge stainless steel tube. A constriction device closes the silastic tube during transfer between reservoirs. To exchange the $<20 \mu l$ volume between pipette tip and reservoir takes ~ 30 –60 s. **B**, Superimposed sample current records taken from **A**, just before (2)(a) and during (3)(b) exposure to ouabain. The pulse potential is indicated for each record; holding potential -40 mV. Records were low-pass-filtered at 2 kHz and digitized at 0.4-ms intervals. **c**, Superimposed difference currents obtained by computer subtraction of each record in **b** from its counterpart in **a**; pulse potentials were, from top to bottom, 0, +40, -80 and -120 mV. Note the expanded current scale in **c**.

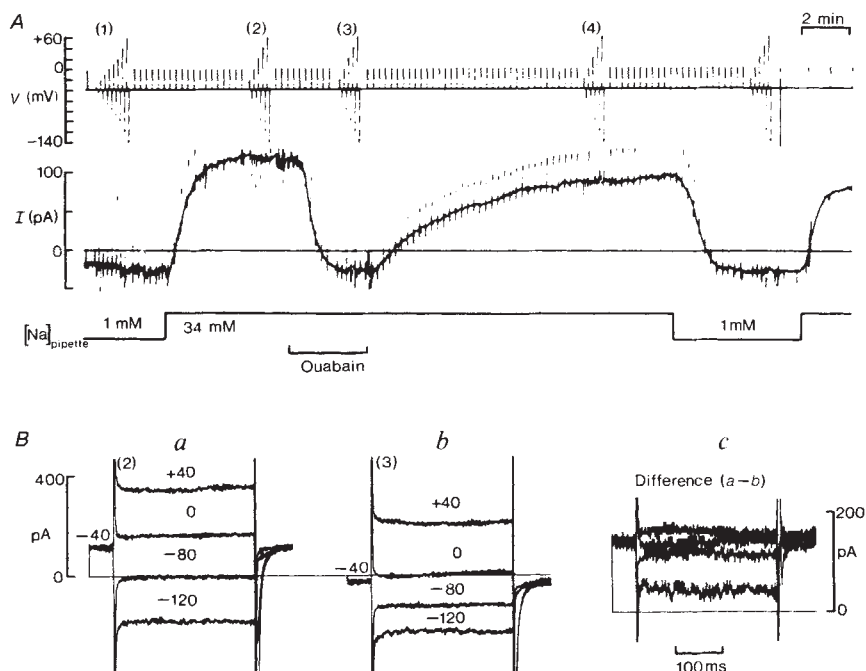
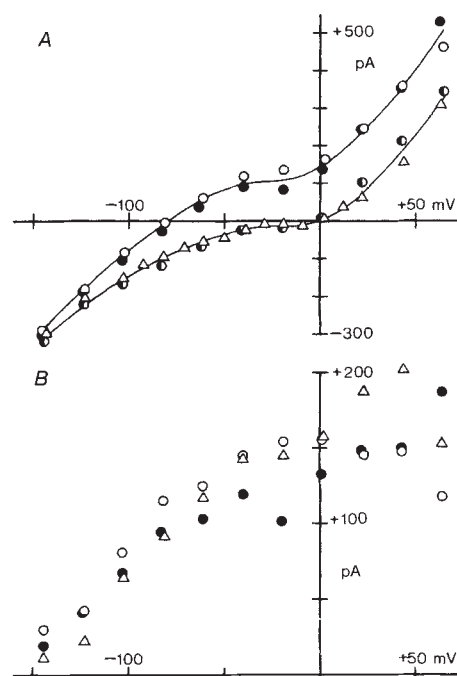


Fig. 2 Voltage dependence of current associated with Na/K pump activity. Data from the experiment in Fig. 1. **A**, Net membrane currents in the presence and absence of Na/K pump activity. Ordinate, steady current levels, obtained by averaging digitized values over the final ~ 100 ms of each 300-ms pulse. Abscissa, membrane potential during the pulse. The data were recorded during the periods numbered in Fig. 1A: Δ (1), 1 mM $[Na^+]_{\text{pipette}}$; $\circ$ (2), 34 mM $[Na^+]_{\text{pipette}}$; $\bullet$ (3), 34 mM $[Na^+]_{\text{pipette}}$ during exposure to ouabain; $\bullet$ (4), 34 mM $[Na^+]_{\text{pipette}}$ after recovery from ouabain. Arbitrary curves were fitted to the points by eye. **B**, Steady-state levels of difference currents derived by subtracting currents recorded in the absence of Na/K pump activity from those recorded, at the same voltage, in its presence. Ordinate, amplitudes measured as in **A**; abscissa, membrane potential. The three sets of points are: Δ (2)–(1), current activated by raising $[Na^+]_{\text{pipette}}$; and ouabain-sensitive current on ouabain application [$\circ$, (2)–(3)] and ouabain withdrawal [$\bullet$, (4)–(3)]. Because these ventricular cells have a flat, rectangular shape, the apparent cell surface was approximated as $2 \times \text{mean width} \times \text{mean length}$. Comparison with cell capacitance measurements, however, suggests that actual membrane area is 2–3 times greater than this approximation, presumably because of T-system and sarcolemmal caveolae. The true pump current density is thus probably 2–3 times smaller than that estimated here.



sensitive current attains a maximal amplitude near 0 mV and, at negative potentials, declines roughly linearly with voltage towards an extrapolated zero-current potential near ~ -160 mV. Very similar current-voltage characteristics were obtained in a total of nine cells ($[Na^+]_{\text{pipette}}$ 34–41 mM) exposed to 5.4 mM K^+ solution to which 10–20 μM ouabain or 50 μM strophanthidin was added to inhibit the Na/K pump; the zero-current potential estimated by linear extrapolation averaged -156 ± 14 mV ($\pm$ s.d.), and the current amplitude at 0 mV was $+160 \pm 69$ pA per cell (equivalent to $3 \pm 1 \mu A \text{ cm}^{-2}$ if related to apparent cell surface, but see Fig. 2 legend).

Before we can conclude that cardiotonic steroid-sensitive

current is simply Na/K pump current, possible contributions from two other sources must be considered. One possibility, that cardiotonic steroids alter membrane current other than via their effect on the Na/K pump, is eliminated by the current-voltage relationships in Fig. 3A, which show that if the pump is first inhibited by exposure to K^+ -free solution, application of 50 μM strophanthidin has no effect. A second possibility is contamination by an indirect effect of pump inhibition. Thus, despite our precautions to minimize K^+ -channel currents, the cardiotonic steroid-sensitive current (and also the Na^+ -activated current) may still contain a component reflecting changes in K^+ concentration in an unstirred extracellular space, caused by the

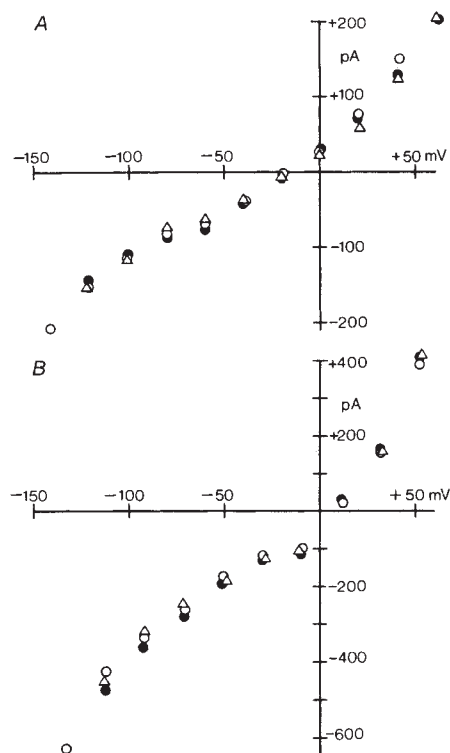


Fig. 3 Control current-voltage relationships demonstrating that, in the absence of Na/K pump activity, membrane current is unaffected either by application of strophanthidin (A), or by a 10-fold change in external $[K^+]$ (B). A, 54 min after raising the $[Na^+]_{\text{pipette}}$ to 41 mM, the Na/K pump was inhibited by exposure to K^+ -free solution, causing a 143-pA reduction in current at 0 mV. Current-voltage relationships were then determined before ($\circ$), during (Δ) and after ($\bullet$) exposure to 50 μM strophanthidin. The points show steady current amplitudes measured as in Fig. 2. Holding potential -40 mV. B, 9 min after raising $[Na^+]_{\text{pipette}}$ to 34 mM, 20 μM ouabain inhibited the Na/K pump, causing a 118-pA inward shift of current at 0 mV. Current-voltage relationships were then determined in 5.4 mM K^+ ($\circ$), 1 mM K^+ (Δ) and 10 mM K^+ ($\bullet$) external solutions, Na^+ replacing K^+ or vice versa. The points show steady currents. Holding potential -30 mV.

changes in pump activity. Figure 3B confirms, however, that once the pump is inhibited (in this case by 20 μM ouabain), a 10-fold change in external $[K^+]$ has virtually no effect on membrane current.

We conclude, therefore, that Fig. 2B illustrates faithfully the Na/K pump current-voltage relationship at high pump rates; its general shape conforms to thermodynamic expectation⁶⁻⁸, as do our preliminary results indicating that a rightward shift of the pump current-voltage relationship accompanies the reduction in pump current caused by lowering either $[Na^+]_{\text{pipette}}$ or external $[K^+]$. Detailed analysis will require higher resolution current measurements to be made, probably using averaging techniques, under various ionic and metabolic conditions.

That such marked voltage dependence of outward Na/K pump current (Fig. 2B) has not been documented previously is probably attributable both to the limited voltage ranges explored (for example, -20 to -90 mV for cardiac preparations^{4,10}), and to the difficulty of clearly distinguishing pump current from other currents. For example, pump-mediated changes in extracellular $[K^+]$ should act via inwardly rectifying K^+ channels to distort current-voltage relations of the type shown in Fig. 2A by making inward K^+ currents smaller at high, than at low, pump rates. This would result in more parallel current-voltage relations than those in Fig. 2A, so that the difference currents (see Fig. 2B) would show less voltage dependence. Voltage sensitivity of inward current generated by the backward-running Na/K pump in squid axons has been reported recently¹⁶, albeit

in the opposite direction to that expected on simple thermodynamic grounds.

The rapid reversal of ouabain action shown here (Fig. 1A; strophanthidin is still severalfold faster) contrasts with its poor reversibility in intact Purkinje fibres from sheep and dog hearts (for example, see ref. 17), but it is not yet clear whether this reflects a mere species difference or results from collagenase treatment of the cells.

Na/K pump current contributes significantly to the electrical activity of cardiac cells^{4,5}; a sustained increase in heart rate, for example, leads to Na/K pump-mediated hyperpolarization and abbreviation of the cardiac action potential^{3,5}. Full analysis or simulation of such effects in heart, nerve or muscle cells requires knowledge of the voltage dependence of the Na/K pump current. Accurate determination of pump current-voltage relationships under various conditions is thus vital to our understanding of both the mechanism of the Na/K pump and its role in modifying electrical activity.

We thank Professors Hiroshi Irisawa and Paul F. Cranefield for encouragement. The work was supported by grants from the USPHS (HL-14899) and the Ministry of Education, Science, and Culture of Japan, and by an Established Fellowship of the New York Heart Association (to D.C.G.).

Received 21 December 1984; accepted 25 February 1985.

1. Thomas, R. C. *Physiol. Rev.* **52**, 563-594 (1972).
2. De Weer, P. in *MTP int. Rev. Sci. Physiol. Ser. 1* Vol. 3 (ed. Hunt, C. C.) 231-278 (Butterworths, London, 1975).
3. Vassalle, M. *Circulation Res.* **27**, 361-377 (1970).
4. Isenberg, G. & Trautwein, W. *Pflügers Arch. ges. Physiol.* **350**, 41-54 (1974).
5. Gadsby, D. C. in *Electrogenic Transport: Fundamental Principles and Physiological Implications* (eds Blaustein, M. P. & Lieberman, M.) 215-238 (Raven, New York, 1984).
6. Hansen, U.-P., Gradmann, D., Sanders, D. & Slayman, C. L. *J. Membrane Biol.* **63**, 165-190 (1981).
7. Chapman, J. B., Johnson, E. A. & Kootsey, J. M. *J. Membrane Biol.* **74**, 139-153 (1983).
8. De Weer, P. in *Electrogenic Transport: Fundamental Principles and Physiological Implications* (eds Blaustein, M. P. & Lieberman, M.) 1-15 (Raven, New York, 1984).
9. Glynn, I. M. in *Electrogenic Transport: Fundamental Principles and Physiological Implications* (eds Blaustein, M. P. & Lieberman, M.) 33-48 (Raven, New York, 1984).
10. Gadsby, D. C. *A. Rev. Biophys. Bioengng* **13**, 373-398 (1984).
11. Hamill, O. P., Marty, A., Neher, E., Sakmann, B. & Sigworth, F. J. *Pflügers Arch. ges. Physiol.* **391**, 85-100 (1981).
12. Matsuda, H. & Noma, A. *J. Physiol., Lond.* **357**, 553-573 (1984).
13. Isenberg, G. & Klockner, U. *Pflügers Arch. ges. Physiol.* **395**, 6-18 (1982).
14. Matsuda, H., Noma, A., Kurachi, Y. & Irisawa, H. *Circulation Res.* **51**, 142-151 (1982).
15. Soejima, M. & Noma, A. *Pflügers Arch. ges. Physiol.* **400**, 424-431 (1984).
16. De Weer, P. & Rakowski, R. F. *Nature* **309**, 450-452 (1984).
17. Deitmer, J. W. & Ellis, D. J. *J. Physiol., Lond.* **284**, 241-259 (1978).

NMR technique for assessing contributions of heavy and light chains to an antibody combining site

Jacob Anglister, Tom Frey & Harden M. McConnell

Stauffer Laboratory for Physical Chemistry, Stanford University, Stanford, California 94305, USA

Nuclear magnetic resonance (NMR) has been used extensively to study the structure of antibody combining sites^{1,2}. In recent studies we have observed the proton resonance spectra of the Fab fragment of a monoclonal anti-spin-label antibody derived from a hybridoma grown on various specifically deuterated amino acids^{3,4}. The broadening of the proton resonance signals by the paramagnetic hapten, together with selective deuteration, has allowed the identification of most of the amino acids in the combining-site region of this antibody and has also provided estimates of distances between amino-acid protons and the unpaired electron⁵. Here we show how recombination of specifically deuterated heavy and light chains permits the assignment of single amino-acid proton resonance signals to either the heavy or light chain. In addition, the spectra of such recombinants demonstrate that their combining-site structures must be almost identical to the native structure.

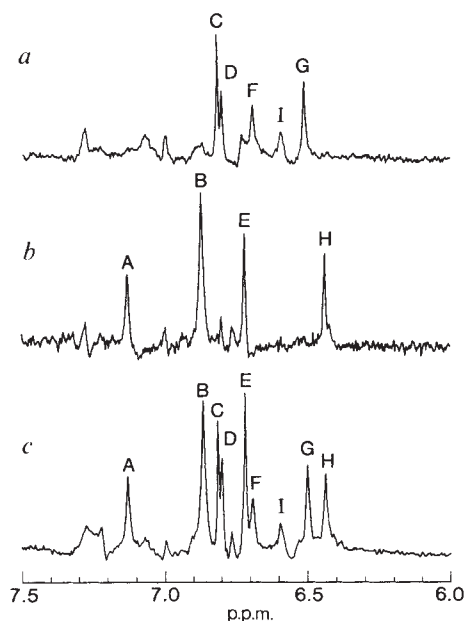
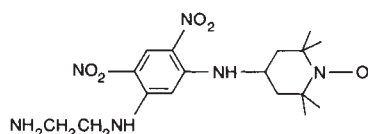


Fig. 1 NMR difference spectra for selectively deuterated Fab fragments of a monoclonal anti-spin-label antibody. *c*, Signals A–I are each due to pairs of protons in the 3,5 ring positions of tyrosine residues in the combining-site region of the ANO2 antibody. Spectra *a* and *b* give the corresponding signals from protons in the light chain and heavy chain, respectively. Note that the spectrum in *c* is, to a good approximation, the sum of the spectra in *a* and *b*. (See text for details.)

The murine IgG₁ monoclonal anti-spin-label antibody (ANO2) was derived from a hybridoma grown on specifically deuterated amino acids, as described previously³. Two preparations of Fab were obtained: one (Fab I) was obtained from the hybridoma grown on perdeuterated tryptophan, perdeuterated phenylalanine and perdeuterated tyrosine; the second (Fab II) was obtained from the hybridoma grown on perdeuterated tryptophan, perdeuterated phenylalanine, and tyrosine deuterated at the 2,6 ring positions. The Fab fragments, prepared as described previously³, were partially reduced and alkylated. The heavy (H)-chain fragment and the light (L) chain were separated on a DEAE column under denaturing conditions (8 M urea, Tris-glycine buffer, pH 8.2), then the heavy-chain fragments and light chains were recombined to produce the recombinants H(I)L(II) and H(II)L(I). The urea solution was made 1 M in propionic acid. The recombinants were dialysed against 1 M propionic acid and then twice against each of the following: water, 5 mM acetate buffer pH 5.8, 5 mM acetate buffer pH 5.8 containing 0.15 M NaCl, and phosphate-buffered saline pH 7.2 (at least 12 h for each). A similar procedure for reconstitution of a whole antibody molecule has been described previously by several groups^{6,7}.

Figure 1 shows the proton resonance difference spectra $\Delta H(II)L(II)$, $\Delta H(II)L(I)$ and $\Delta H(I)L(II)$, where Δ = proton resonance spectrum of Fab – proton resonance spectrum of the Fab fragment with paramagnetic spin-label hapten bound. The paramagnetic spin-label hapten has the formula



The proton resonance signals in Fig. 1*a* arise from the 3,5 protons of tyrosines in the light chain that are in the combining-site region (within 17 Å of the unpaired spin⁵). The signals in

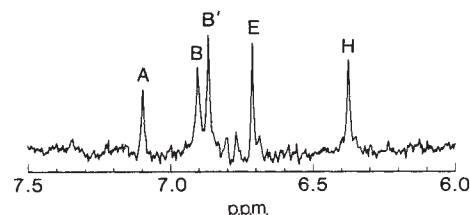


Fig. 2 NMR difference spectrum for a selectively deuterated Fab fragment of a monoclonal anti-spin-label antibody. The proton resonance signals arise from pairs of protons in the 3,5 positions of the tyrosine residues of the Fab heavy chain. These resonance signals are the same as those in Fig. 1*b*, except that in the present case the combining site is occupied by the chemically reduced non-paramagnetic form of the spin-label hapten (the hydroxylamine) whereas the signals in Fig. 1*b* arise from a combining site that is unoccupied. Note that the presence of the hapten leads to splitting of signal B in Fig. 1*b* (due to two tyrosines and four protons) into two signals (B and B') in Fig. 2.

Fig. 1*b* are due to the 3,5 protons of the tyrosines of the heavy chain in the combining-site region.

That the spectrum in Fig. 1*c* is very nearly the sum of the spectra in Fig. 1*a* and *b* represents compelling evidence that the combining-site structures are virtually the same in all three cases; this is because resonance positions, intensities and linewidths are sensitive to conformation, motional freedom and environment. A similarity in structure between native and reconstituted IgG was observed previously using circular dichroism and optical rotatory dispersion^{7,8}.

Our spectra show that there are five tyrosines in each chain in the combining-site region. Signal I from the light chain is weaker because it is farther from the unpaired electron. In each tyrosine signal the two proton resonance lines are equivalent due to rapid rotation of the aromatic ring. Similar conclusions are reached from the proton resonance spectra of the Fab fragment containing tyrosine deuterated in the 3,5 positions (data not shown), that is, there are five tyrosines in each chain, and the conformation of the combining site in this reconstituted Fab is very similar to the conformation in the native Fab.

Signal B in Fig. 1*b* arises from two tyrosines, as is evident from the observed intensity. Figure 2 shows that this signal is split in a difference spectrum, Δ = spectrum of Fab containing the reduced spin label (nitroxide to hydroxylamine) – resonance spectrum of Fab to which the paramagnetic spin label is bound. Interestingly, all the tyrosine proton resonance signals in Fig. 2 remain sharp, and some shift in resonance position when the non-paramagnetic hapten is in the combining site. Presumably, some of these tyrosine residues are in contact and/or are hydrogen-bonded to the hapten. Nonetheless, the motional freedom remains large enough to yield sharp resonance signals (4–7 Hz linewidth at half height).

In previous studies using double difference spectra, we demonstrated the presence of three to five threonines in the combining-site region⁴. In the present study we have used double difference spectra involving two recombinant Fab fragments to establish the presence of two threonine residues on the light chain and two to three threonine residues on the heavy chain in the combining-site region. One Fab fragment was not labelled and the other had deuterated threonine in the light chain (data not shown).

Note that the 'combining-site region' is defined as the region within 17 Å of the unpaired electron, and that proton resonance signals in this region are obtained from the difference spectra Δ = proton resonance spectrum of Fab – proton resonance spectrum of Fab with paramagnetic spin-label hapten bound. The geometry of the hapten is such that the protons of all amino acids in contact with the hapten are included in this combining-site region. The 'combining site' is thus fully included within

the 'combining-site region.' Integration of the difference spectra and assignment of the resonances to specific types of amino acids reveal that the difference spectra represent contributions from ~50 amino acids^{3,4}, while the entire Fab fragment contains 440 amino acids. It is probable that some, or all, of the tyrosine residues have a significant role in affecting the structure of the combining site, irrespective of whether they are in contact with the hapten or not.

The present study demonstrates that NMR, coupled with specific amino-acid deuteration and with heavy- and light-chain recombination, can be used to establish the contributions of the heavy and light-chain amino acids of an IgG monoclonal antibody to the combining-site region, provided the relevant hapten or antigen produces a suitable perturbation of nuclear resonance signals. Therefore, NMR studies offer the possibility of obtaining a working model of the combining-site structure of the ANO2 monoclonal antibody when the complete amino-acid sequence becomes available. The difference between the spectra in Fig. 1b and Fig. 2 indicates the information potentially available for non-paramagnetic haptens/antigens. Some of the resonances in the spectrum in Fig. 2 undergo changes in chemical shift relative to their positions in the spectrum in Fig. 1b. These resonances show up in a difference spectrum; Δ = spectrum with reduced spin label - spectrum without hapten, implying that useful information on combining-site structure may sometimes be obtained even with non-paramagnetic haptens.

This work was supported by ONR Contract N00014-83-K-0349 (to H.M.M.) and American Cancer Society Fellowship S2-84, California Division (to J.A.). We thank Dr Norman Klinman for advice on the recombination experiments.

Received 7 November 1984; accepted 13 February 1985.

1. Dower, S. & Dwek, R. A. in *Biological Application of Magnetic Resonance* (ed. Shulman, R. G.) 271-303 (Academic, New York, 1979).
2. Goetze, A. M. & Richards, J. H. *Biochemistry* 17, 1733-1739 (1978).
3. Anglister, J., Frey, T. & McConnell, H. M. *Biochemistry* 23, 1138-1142 (1984).
4. Frey, T., Anglister, J. & McConnell, H. M. *Biochemistry* 23, 6470-6473 (1984).
5. Anglister, J., Frey, T. & McConnell, H. M. *Biochemistry* 23, 5372-5375 (1984).
6. Metzger, H. & Mannik, M. *J. exp. Med.* 120, 765-782 (1964).
7. Bjork, I. & Tanford, C. *Biochemistry* 10, 1289-1295 (1971).
8. Stevenson, G. T. & Dorrington, K. J. *Biochem. J.* 118, 703-712 (1970).

Comparison of α -tropomyosin sequences from smooth and striated muscle

Nelson Ruiz-Opazo, Judah Weinberger*
& Bernardo Nadal-Ginard†

Laboratory of Molecular and Cellular Cardiology, Department of Cardiology, Children's Hospital and Department of Pediatrics, Harvard Medical School, 300 Longwood Avenue, Boston, Massachusetts 02115, USA

Tropomyosins are a closely related family of proteins with a dimeric α -coiled-coil structure¹. Skeletal isoforms are composed of two types of subunits, α and β ²⁻⁴ which, in turn, are assorted into two main molecular species $\alpha\alpha$ and $\alpha\beta$ ⁵⁻⁷. Both isoforms are present in different molar ratios in individual skeletal muscle types^{8,9}. In small mammals, however, only α -chain is expressed in cardiac muscle⁸. Tropomyosin, in association with the troponin complex (troponin-I, -T and -C) plays a central role in the Ca^{2+} -dependent regulation of vertebrate striated muscle contraction¹⁰⁻¹². On the other hand, despite structural similarities with the striated isoforms, the function of this protein in smooth muscle and non-muscle cells remains unknown, because in these cells contraction is thought to be regulated by myosin-linked processes independently of tropomyosin¹³⁻¹⁵. Here we report the nucleotide

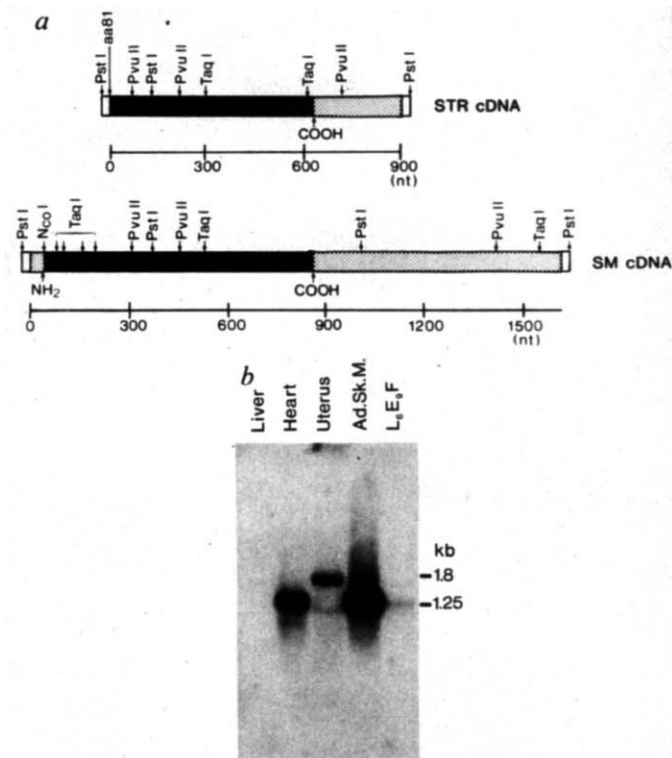


Fig. 1 Isolation and characterization of rat smooth and striated muscle α -tropomyosin cDNA clones. *a*, Partial restriction endonuclease map of striated (STR) and smooth (SM) muscle cDNA clones. Solid, stippled and open bars indicate amino-acid coding, 3' and 5' untranslated regions and GC tails, respectively. The position of the first (residue 81, -NH₂) and last (-COOH) codons are indicated. The uterus-derived cDNA clone (~1.7 kb) contains mRNA sequences starting at the poly(A) track and extends all the way through the 5' untranslated region. The striated muscle-derived cDNA clone (~0.9 kb) spans the mRNA from the poly(A) track up to codon 81. *b*, Hybridization of the smooth muscle cDNA clone to a panel of total cellular RNAs (20 μ g per lane) isolated from different tissues. The ³²P-labelling (nick-translation) hybridization and blotting procedures were performed as described previously⁴⁰. The size of the two hybridizing mRNA species is indicated in kb. Ad. Sk. M. and L₆E₉F represent hind leg adult skeletal muscle and L₆E₉ myotubes¹⁶, respectively. The cloning of striated and smooth muscle cDNA was performed with mRNA-DNA hybrids annealed to pBR322 essentially as described^{41,42} using RNA isolated by the hot phenol procedure⁴³ from adult rat hind leg and gravid uterus. One α -tropomyosin cDNA clone was isolated from each cDNA library using a previously characterized cDNA clone which covers residues 80-120 in the α -tropomyosin sequence⁴⁴.

sequences of cloned complementary DNAs for rat striated and smooth muscle α -tropomyosin. Comparison of the derived amino-acid sequences reveals the existence of tissue-specific peptides that delimit the putative troponin-I and troponin-T binding domains of tropomyosin. S₁-nuclease mapping studies reveal the existence of three distinct α -tropomyosin messenger RNA isoforms each encoding a different protein; these isoforms are tissue-specific, developmentally regulated and most probably encoded by the same gene.

Partial restriction maps of smooth and striated muscle α -tropomyosin cDNA clones are shown in Fig. 1a. The uterus-derived cDNA clone, of some 1.7 kilobases (kb), contains mRNA sequences starting at the poly(A) track and extends all the way through the 5'-untranslated region. The striated muscle-derived cDNA clone (~0.9 kb) spans the mRNA from the poly(A) track up to residue 80. To ascertain the tissue-specificity of the two cDNA clones, we performed blot analysis of RNA isolated from different tissues (Fig. 1b). Using the smooth muscle cDNA clone as probe, two size-classes of α -tropomyosin mRNA were detected: a 1.25-kb mRNA present in adult ventricular

* Present address: Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts 02115, USA.

† To whom correspondence should be addressed.

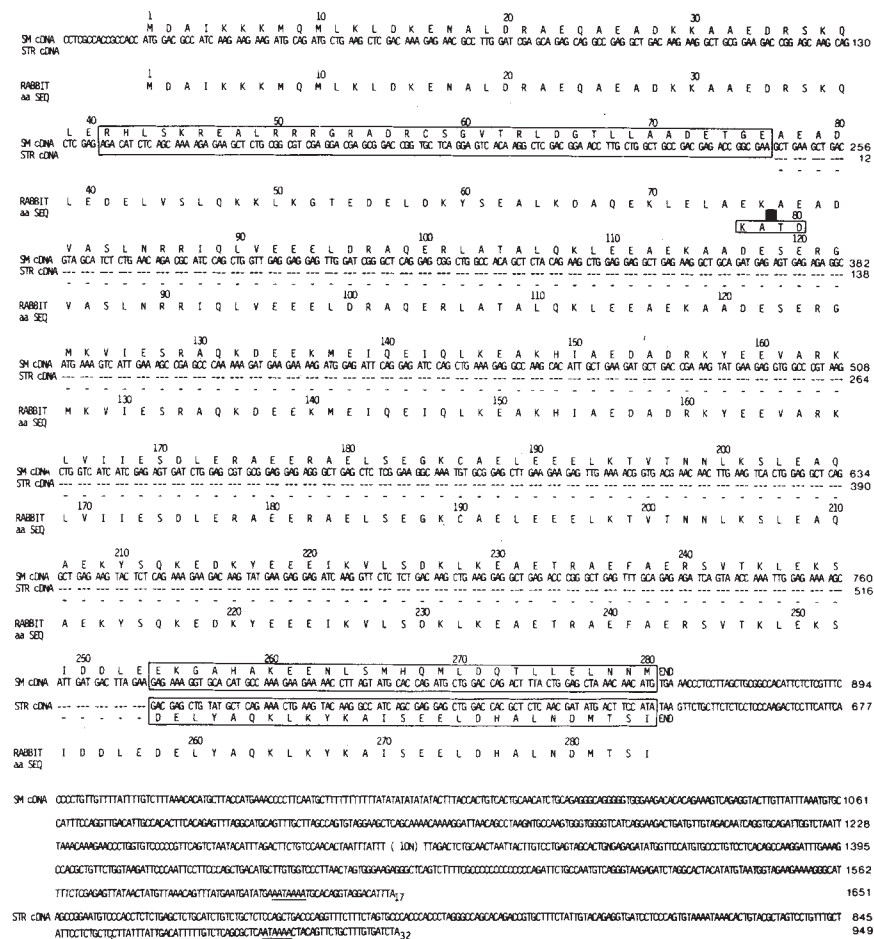


Fig. 2 Nucleotide and derived amino-acid sequence comparison of smooth and striated muscle α -tropomyosin cDNA clones. Nucleotide sequence of the smooth and striated muscle α -tropomyosin-cDNA clones were determined by cloning of gel-purified restriction fragments from both cDNA clones into M13 mp cloning vectors and subsequently sequenced according to Sanger *et al.*⁴⁵. The deduced amino-acid sequences for the smooth and striated muscle α -tropomyosins are shown on top and underneath, respectively, of the corresponding nucleotide sequences. The AATATAA sequence of both cDNAs are underlined. Note that a second AATATAA sequence is located 75 nucleotides upstream of the used one in the striated muscle cDNA sequence, suggesting the existence of two different poly(A) addition sites. The isotype-switch peptides sequences are boxed (residues 41–76 and 254–280 of the smooth muscle isoform). Most of the striated muscle cDNA sequence (residues 77–253, smooth muscle isoform) is shown as dotted lines indicating identity with the analogous region of the smooth muscle isoform. The rabbit α -tropomyosin residue sequence³ has been aligned with the rat α -tropomyosin isoforms by boxing the extra four amino acids out of the linear arrangement of the sequence.

myocardium, uterus, skeletal muscle and L₆E₉ myotubes¹⁶ (L₆E₉F) and a 1.8-kb-long mRNA expressed mainly in uterus. Uterine cells contain the highest concentration of the 1.8-kb mRNA and the lowest of the 1.25-kb species, whereas the converse is true for heart and skeletal muscle. Longer exposure of the L₆E₉F lane also shows that, although both mRNAs are present in the myotubes of this clonal cell line, the 1.25 kb species is predominant. Based on the differences in size and tissue distribution, the 1.25- and 1.8 kb mRNA species probably represent the mRNA coding for the striated and smooth muscle α -tropomyosin mRNA isoforms, respectively.

Examination of the nucleotide sequences shown in Fig. 2 reveals several new structural features of tropomyosin proteins. First, the smooth muscle α -tropomyosin is 280-residues long and 4 residues shorter than the rabbit skeletal muscle isoform³. Second, the rat striated muscle isoform is identical in amino-acid sequence to its rabbit counterpart, at least throughout the regions which have been compared (residues 81–284 of the rabbit molecule). Third, the comparative analysis between the smooth and striated muscle isoforms uncovers 'isotype-switch peptides' that reside in different positions along the protein molecule. The striated muscle α -tropomyosin is identical in amino-acid sequence to the smooth muscle isoform except for residues 258–284 (striated muscle sequence) that defines a carboxy-terminal 'isotype-switch peptide'. In addition, comparison of the rabbit striated muscle α -tropomyosin with the rat smooth muscle isoform shows them to be identical in sequence except for two regions: from residues 41–80 (striated muscle sequence), which depicts an amino-proximal isotype-switch peptide, and the carboxy-terminal isotype-switch peptide discussed above. By S₁ nuclease mapping, we have demonstrated that the rat striated muscle α -tropomyosin also has the amino-proximal isotype switch peptide in common with rabbit striated muscle

(data not shown). Further inspection of the smooth muscle α -tropomyosin amino-acid sequence in the amino-proximal isotype-switch peptide (smooth muscle residues 41–76) reveals that the 3–4–3–4-periodicity of hydrophobic residues characteristic of the tropomyosin proteins^{17,18} is interrupted significantly at positions 46 (R), 50 (R), 53 (G), 57 (R) and 60 (G) (Fig. 2). The occurrence of such residues, especially those that are charged, could lead to a local disruption of the α -helix, which has been suggested to affect the stability of the coiled-coil at the protein level¹⁹. These local differences in structure could alter the interaction of the smooth muscle α -tropomyosin with other components of the contractile apparatus.

Recently, it has been suggested that the binding of troponin-I to tropomyosin enhances cooperatively the binding of troponin-T to tropomyosin²⁰; however, the location of the troponin-I binding site on tropomyosin has not yet been determined. The interaction of tropomyosin with troponin-T seems to occur near Cys 190 (site 2)^{21–27} and at a site close to or at the carboxy-terminus (site 1) of the tropomyosin molecule^{28,29}. As shown in Fig. 2, the only difference in primary sequence between the smooth and striated α -tropomyosins resides on the isotype-switch peptides. Therefore, it is likely that these two peptides specify the troponin-I (residues 41–80) and troponin-T site 1 (residues 258–284) binding domains, respectively. These isotype-specific peptides also can explain the observed differences in actin binding and in the actin activation of myosin between the two α -tropomyosin isoforms³⁰. At the nucleotide level, the most striking observation from the comparative analysis between the striated and smooth muscle cDNA clones is their identity in the region coding for residues 81–257 (striated muscle sequence) (Fig. 2). In contrast, sequences located 3' from codon 257 are unique for each cDNA. The fact that in the common amino-acid coding sequence of the two mRNAs there is not a

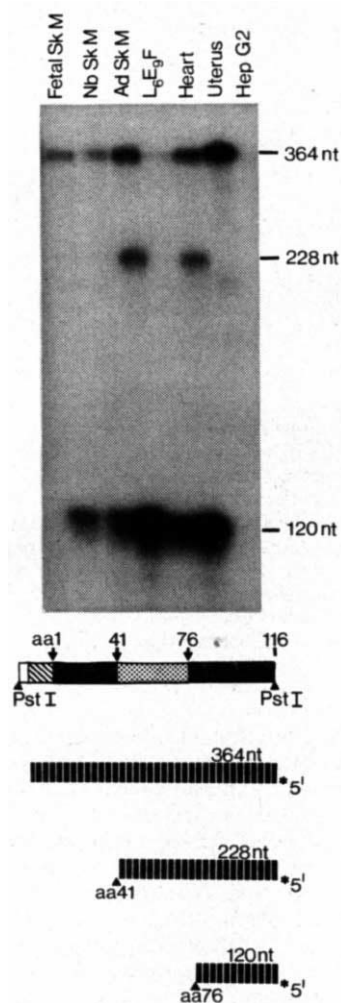


Fig. 3 Developmental and tissue-specific expression of α -tropomyosin mRNAs. The S_1 nuclease mapping studies were done under condition essentially described by Mahdavi *et al.*⁴⁶. The 5' end-labelled fragment was hybridized as a double-stranded probe to 30 μ g total RNA under RNA-loop conditions^{47,48}. After hybridization, the mixture was incubated with 600 U m^{-1} of enzyme and the S_1 nuclease-resistant products were analysed by 6% polyacrylamide/8.3 M urea gel electrophoresis⁴⁶ and their sizes estimated by comparison with OX174 size standards. RNA samples isolated from 18-day-old fetal skeletal muscle (fetal Sk.M.), newborn skeletal muscle (Nb.Sk.M.), adult skeletal muscle (Ad.Sk.M.), adult ventricular myocardium (heart), gravid uterus (uterus), L_6E_9 myotubes (L_6E_9F) and hepatocellular carcinoma cell line (HepG₂) as a negative control were analysed. The 5' end-labelled probe was generated by end-labelling of the restriction fragment with [γ -³²P]ATP (NEN; specific activity 3,000 Ci $mmol^{-1}$) using T_4 polynucleotide kinase (Boehringer Mannheim). The 364-base pair long 5'-most $PstI/PstI$ restriction fragment from the smooth muscle cDNA clone (Fig. 1a), labelled at the 5' end with [γ -³²P], was used in the S_1 nuclease mapping studies. This probe contains the 16 nucleotides of the 5' untranslated region of the smooth muscle cDNA down to codon 116, thus including the sequence coding for the amino-proximal isotype-switch peptide (smooth muscle sequence residues 41–76). The $PstI$ site at codon 116 lies in a region which is common to both types of cDNAs. Therefore, protection from S_1 nuclease digestion of the full-length probe (364 bases) indicated hybridization to homologous smooth muscle α -tropomyosin mRNA, whereas protection of a 120-base-long fragment represents hybridization to the striated muscle α -tropomyosin mRNA. Protection of subfragments other than 120 bases long would indicate the existence of new α -tropomyosin isoforms that diverge from the smooth muscle sequence at positions different from the striated muscle isoform.

single nucleotide substitution (even at the wobble position) strongly suggests that a single gene specifies the expression of these two tissue-specific α -tropomyosin isoforms by a complex process of alternative splicing. This possibility has been confirmed recently by the isolation of a single α -tropomyosin gene that contains all the common and isotype-specific sequences of the cDNA clones (N.R.-O. and B.N.-G., in preparation).

To study the developmental and tissue-specific expression of the α -tropomyosin mRNAs, S_1 -nuclease mapping analysis of RNA isolated from different rat tissues was performed as described in Fig. 3 legend. Three distinct fragments are protected from S_1 -nuclease digestion by skeletal and ventricular adult muscle RNA. Two of these fragments, 120- and 364-bases long, are consistent with the sizes predicted to be protected by striated (α_1 tropomyosin) and smooth muscle (α -tropomyosin) mRNAs, respectively. The third band, 228-bases long, demonstrates the existence of a new α -tropomyosin mRNA (striated muscle α_2 tropomyosin) that diverges from the smooth muscle cDNA probe at the region corresponding to codon 41, which is the first residue of the amino-proximal isotype-switch peptide (smooth muscle residues 41–76). Identity between the striated muscle α_2 - and smooth muscle α -tropomyosin mRNAs spans the region encoding residues 41–116 and, therefore, includes the amino-proximal isotype-switch peptide (Fig. 3).

The three different α -tropomyosin mRNAs are tissue-specific and developmentally regulated. The α_1 - and α_2 -tropomyosin mRNAs are produced only in striated muscle, whereas the α -form seems to be expressed predominantly in smooth muscle. The presence of low levels of the smooth muscle α -tropomyosin isoform in striated muscle indicates either tissue contamination from the vascular smooth muscle or a low level of expression of the smooth muscle isoform in the striated muscle cells. The fact that L_6E_9 myotubes produce both α_1 - and α -tropomyosin mRNA favours the second possibility and demonstrates that striated and smooth muscle isoforms can be synthesized in the same cell. The striated muscle α -tropomyosin isoforms are also developmentally regulated. The α_1 -tropomyosin mRNA is induced early in muscle development and accumulates to high levels in the adult, whereas α_2 -tropomyosin mRNA can only be detected in adult skeletal and ventricular tissues (Fig. 3). These results also demonstrate that identical α -tropomyosin isoforms are expressed in skeletal and cardiac muscles, despite the fact that the two muscle types express different myosin heavy chains^{31–33}, light chains^{34,35}, actin³⁶ and troponins^{37–39}.

The results presented here show the existence of three distinct α -tropomyosin mRNAs, all three of which share regions of 100% homology and isotype specific sequences. Such an arrangement suggests that the α -tropomyosin gene has an unusual organization with common and exchangeable exons that can be spliced together according to different combinatorial possibilities.

We thank Drs S. Hitchcock-DeGregori, S. Grund and V. de Leon for critical reading of the manuscript, S. Movahedi for technical assistance, M.-A. Strehler for computer analysis and M. Hagar, M. Hawkins and B. Trudeau for secretarial assistance. This research was supported by grants from NIH, American Heart Association, American Cancer Society and Muscular Dystrophy Association.

Received 27 September 1984; accepted 25 February 1985.

- Smillie, L. B. *Trends biochem. Sci.* **4**, 151–154 (1979).
- Sodek, J., Hodges, R. S. & Smillie, L. B. *J. biol. Chem.* **253**, 1129–1136 (1978).
- Stone, D. & Smillie, L. B. *J. biol. Chem.* **253**, 1137–1148 (1978).
- Mak, A. S., Lewis, W. G. & Smillie, L. B. *FEBS Lett.* **105**, 232–234 (1979).
- Eisenberg, E. & Kielley, W. W. *J. biol. Chem.* **249**, 4742–4748 (1974).
- Yamaguchi, M., Greaser, M. & Cassens, R. G. *J. ultrastruct. Res.* **48**, 33–48 (1974).
- Lehrer, S. S. *Proc. natn. Acad. Sci. U.S.A.* **72**, 3377–3381 (1975).
- Cummins, P. & Perry, S. V. *Biochem. J.* **133**, 765–777 (1973).
- Cummins, P. & Perry, S. V. *Biochem. J.* **43**, 49 (1974).
- Perry, S. V. *Biochem. Soc. Trans.* **7**, 593 (1979).
- Taylor, E. W. *CRC Crit. Rev. Biochem.* **6**, 103–164 (1979).
- Adelstein, R. S. & Eisenberg, E. A. *Rev. Biochem.* **49**, 921–956 (1980).
- Gorecka, A., Aksoy, M. O. & Hartshorne, D. J. *Biochem. biophys. Res. Commun.* **71**, 325–331 (1976).
- Sobieszek, A. & Small, J. V. *J. molec. Biol.* **101**, 75–92 (1976).

15. Chacko, S., Conti, M. A. & Adelstein, R. S. *Proc. natn. Acad. Sci. U.S.A.* **74**, 129-133 (1977).
16. Nadal-Ginard, B. *Cell* **15**, 855-864 (1978).
17. Elzinga, M. & Trus, B. in *Methods in Peptide and Protein Sequence Analysis* (ed. Birr, C.) 213-224 (Elsevier, Amsterdam, 1980).
18. Stone, D., Sodek, J., Johnson, P. & Smillie, L. B. *Proc. IX FEBS Meet.* **31**, 125-136 (1975).
19. Lehrer, S. S., Betteridge, D. R., Graceffa, P., Wong, S. & Seidel, J. C. *Biochemistry* **23**, 1591-1595 (1984).
20. Pearlstone, J. & Smillie, L. B. *J. biol. Chem.* **258**, 2534-2542 (1983).
21. Nonomura, Y., Drabikowski, W. & Ebashi, S. *J. Biochem., Tokyo* **64**, 419-422 (1968).
22. Cohen, C., Caspar, D. L. D., Parry, D. A. D. & Lucas, R. M. *Cold Spring Harb. Symp. quant. Biol.* **36**, 205-215 (1971).
23. Margossian, S. S. & Cohen, C. *J. molec. Biol.* **81**, 409-413 (1973).
24. Ohtsuki, I. *J. Biochem., Tokyo* **75**, 753-765 (1974).
25. Stewart, M. & McLachlan, A. D. *J. molec. Biol.* **103**, 251-260 (1976).
26. Morris, E. P. & Lehrer, S. S. *Biophys. J.* **33**, 239a (1981).
27. Chong, P. C. S. & Hodges, R. S. *J. biol. Chem.* **257**, 9152-9160 (1982).
28. Pato, M. D., Mak, A. S. & Smillie, L. B. *J. biol. Chem.* **256**, 602-607 (1981).
29. Mak, A. S. & Smillie, L. B. *J. molec. Biol.* **149**, 541-550 (1981).
30. Yamaguchi, M., Ver, A., Carlos, A. & Seidel, J. C. *Biochemistry* **23**, 774-779 (1984).
31. Mahdavi, V., Periasamy, M. & Nadal-Ginard, B. *Nature* **297**, 659-665 (1982).
32. Nguyen, H. T., Wydro, R. M., Gubits, R. M. & Nadal-Ginard, B. *Proc. natn. Acad. Sci. U.S.A.* **79**, 5230-5234 (1982).
33. Wydro, R. M., Nguyen, H. T., Gubits, R. M. & Nadal-Ginard, B. *J. biol. Chem.* **258**, 670-678 (1983).
34. Lowey, S. & Risby, D. *Nature* **234**, 81-85 (1971).
35. Gauthier, G. F., Lowey, S. & Hobbs, A. W. *Nature* **274**, 24-29 (1978).
36. Nudel, U. *et al. Proc. natn. Acad. Sci. U.S.A.* **79**, 2763-2767 (1982).
37. Dhoot, G. K. & Perry, S. V. *Nature* **278**, 714-718 (1979).
38. Toyota, N. & Shimada, Y. *J. Cell Biol.* **91**, 497-504 (1981).
39. Toyota, N. & Shimada, Y. *Cell* **33**, 297-304 (1983).
40. Chakraborty, P. R., Ruiz-Opazo, N., Shouval, D. & Shafritz, D. A. *Nature* **286**, 531-533 (1980).
41. Wood, K. O. & Lee, J. C. *Nucleic Acids Res.* **3**, 1961-1971 (1976).
42. Zain, S., Sambrook, J., Roberts, R. J., Keller, W., Fried, M. & Dunn, A. R. *Cell* **16**, 851-861 (1979).
43. Soeiro, R., Birnboim, H. C. & Darnell, J. E. *J. molec. biol.* **19**, 362-372 (1966).
44. Garfinkel, L. I., Periasamy, M. & Nadal-Ginard, B. *J. biol. Chem.* **257**, 11078-11086 (1982).
45. Sanger, F., Nichlen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463-5467 (1977).
46. Mahdavi, V., Chambers, A. P. & Nadal-Ginard, B. *Proc. Natn. Acad. Sci. U.S.A.* **81**, 2626-2630 (1984).
47. Casey, J. & Davidson, N. *Nucleic Acids Res.* **4**, 1539 (1977).
48. Berk, A. J. & Sharp, P. A. *Cell* **12**, 721 (1977).

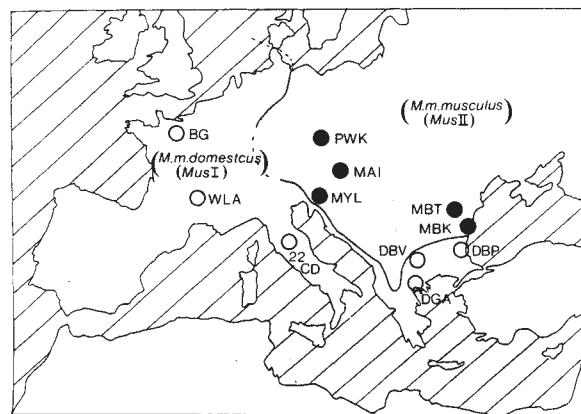


Fig. 1 Map of Europe showing the geographical origin of the 10 wild-derived strains of mouse studied. The two semispecies meet and hybridize along a narrow zone, indicated by the thick line, extending from Denmark, through central Europe to the Black Sea^{2,4,5}. *M.m. domesticus* is found in western and southern Europe, *M.m. musculus* in central and eastern Europe. These semispecies have also been termed *Mus 1* and *Mus 2* (see ref. 5). The distribution of the two Y-chromosome haplotypes of these strains (indicated by solid and open circles) shows that each is specific to one of the semispecies.

Most classical *Mus musculus domesticus* laboratory mouse strains carry a *Mus musculus musculus* Y chromosome

C. E. Bishop*, P. Boursot†, B. Baron‡, F. Bonhomme† & D. Hatat*

* Unité d'Immunogénétique Humaine, INSERM U276, and

‡ Unité de Biochimie des Régulations Cellulaires, Service Commun INSERM-Pasteur (Trieur de Cellules), Institut Pasteur, 75724 Paris, Cedex 15, France

† Institut des Sciences de l'Evolution, Université de Montpellier II, 34060 Montpellier, France

Restriction fragment length polymorphisms (RFLPs) of maternally inherited mitochondrial DNA have revealed that genetic exchanges can occur between populations, subspecies or even species¹⁻⁴. From the point of view of population genetics the mammalian Y chromosome represents a genomic analogue of the mitochondrion; it is inherited only paternally and remains perpetually monosomic, showing little meiotic recombination with other chromosomes. Using a Y-specific genomic DNA probe obtained from a flow-sorted mouse Y-chromosome library, we have examined the RFLPs in 10 newly established mouse lines of the European semispecies *Mus musculus domesticus* and *Mus musculus musculus*⁵, and identified two variant forms of the Y chromosome, each of which is characteristic of one of the semispecies. As reported here, probing the DNA from nine established inbred laboratory strains reveals that the strain SJL carries a *M.m. domesticus* type Y whereas, surprisingly, the Y chromosomes of the other eight strains are of *M.m. musculus* origin. Hence, these strains cannot be regarded as archetypes of *M.m. domesticus* as suggested by protein polymorphisms and mitochondrial DNA sequence data, but rather as genetic hybrids between the two semispecies.

Analysis of genetic variation in wild mouse populations affecting either nuclear genes (measured by protein polymorphisms) or maternally inherited cytoplasmic DNA (that is, mitochondrial DNA RFLPs) has revealed the existence of extensive polymorphism in most systems. These differential patterns allow a clear biochemical distinction to be made between different taxonomical units and show that a variety of interrelationships exist between the different groups⁵. In contrast, an analysis of the RFLPs of mitochondrial DNA in common inbred laboratory strains shows little variation, suggesting that these strains may in fact be derived from a single *M.m. domesticus* female lineage⁶. This implies that the extensive nuclear variation which exists, as evidenced by protein polymorphisms, could have been introduced through the males. One way to assess this contribution is to examine RFLPs of the mouse Y chromosome; this chromosome is unique in that normally it can be transmitted only via the male, it remains haploid in the genome and shows only a very limited amount of meiotic recombination with other chromosomes⁷. Polymorphisms characteristic of one population would therefore tend to become fixed on the Y chromosome.

We have shown recently that Y-specific DNA probes can be obtained from Y chromosome-enriched DNA libraries⁸⁻¹⁰. We have cloned two independent complementary DNA libraries from mouse metaphase chromosomes enriched for the Y chromosome by flow sorting according to the methods described by Baron *et al.*¹¹. A random probe (pY353/B) consisting of a 1.5-kilobase (kb) fragment was isolated from one of these libraries; when used to probe *EcoRI*-restricted genomic blots from inbred male and female C57BL/6 mice under either stringent or non-stringent conditions, pY353/B hybridized with the male alone. In addition to the cognate 1.5-kb band, the probe detects four homologous bands of ~5, 7, 9.5 and 15 kb and therefore forms part of a small multi-sequence family. pY353/B reacts with the male teratocarcinoma cell line PCC7 but not with the male teratocarcinoma line PCC4 which has lost the Y chromosome during culture (data not shown). Therefore, we consider this probe to be Y chromosome-specific.

To assess the Y-chromosome polymorphisms that exist in the wild, we used a panel of DNAs from 10 newly established inbred strains; these lines were derived from mice caught at the locations shown in Fig. 1 and subsequently inbred for a minimum of seven generations. The mice fall into the two well established European groups *M.m. domesticus* and *M.m. musculus*. These groups can be distinguished taxonomically or

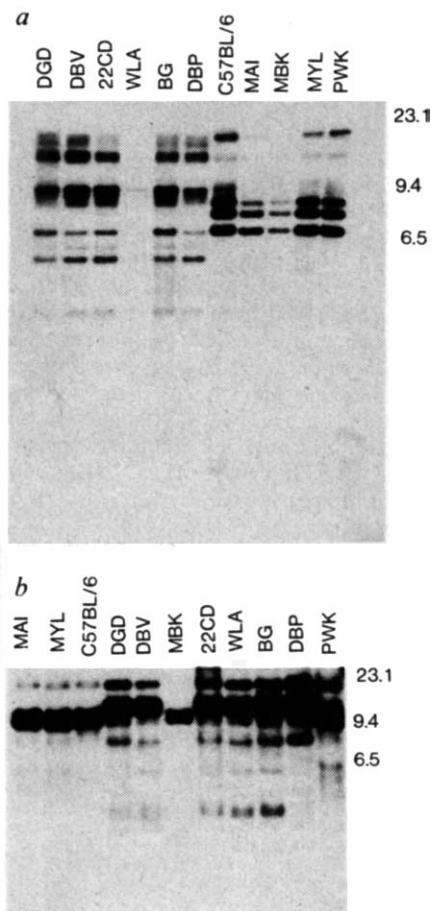


Fig. 2 Hybridization of pY353/B to DNAs from male mice. DNA was extracted from individual male mice and digested with either *MspI* or *HindIII*. DNA (10 µg) was then separated on either 0.7% (a) or 0.8% (b) agarose gels and transferred to a Zetapore filter membrane. pY353/B was labelled to an approximate specific activity of 2×10^8 c.p.m. μg^{-1} with ^{32}P by nick-translation and used to probe the blots. Hybridization was at 42 °C for 16 h in 50% formamide, $5 \times \text{SSC}$. Blots were washed three times for 30 min each at 68 °C in $2 \times \text{SSC}$, 0.1% SDS, followed by three times for 30 min each at 68 °C in $0.1 \times \text{SSC}$, 0.1% SDS. Filters were then exposed to Kodak X-AR5 film for 16–20 h at –70 °C. Approximate sizes in kilobase (kb) are shown at the right. Mouse lines DGD, DBV, 22CD, WLA, BG and DBP are of *M.m. domesticus* type; lines MAI, MBK, MYL and PWK of *M.m. musculus* type. a (*MspI* digestion) shows that laboratory strain C57BL/6 exhibits three bands of ~6.6, 7.0 and 8.0 kb characteristic of *M.m. musculus*. b (*HindIII* digest) again shows that C57BL/6 exhibits the *M.m. musculus* characteristic 10- and 12-kb bands. Re-probing the blots, after dehybridization, with a cloned mouse *H-2* cDNA probe showed that the lanes do not contain exactly equal amounts of DNA. The weaker signal seen with WLA results from underloading.

more recently on the basis of multiple biochemical markers⁵; they are parapatric (*M.m. domesticus* is found in western and southern Europe, *M.m. musculus* in central and eastern Europe) and meet along a narrow hybridization zone^{2,4}. We believe the two groups represent semispecies within the complex *Mus musculus* rather than subspecies, as important restrictions to genetic introgression reduce the hybrid zone to a narrow geographical area; this prevents the formation of a genetic continuum characteristic of subspecies interaction.

DNAs from single male and female mice were digested with *EcoRI*, *TaqI*, *MspI* or *HindIII*, separated on agarose gels and transferred to a Zetapore membrane. pY353/B was labelled with ^{32}P by nick-translation and used to probe the genomic blots. After washing at 68 °C in $0.1 \times \text{SSC}$ we detected no hybridization with any of the female DNAs (data not shown)

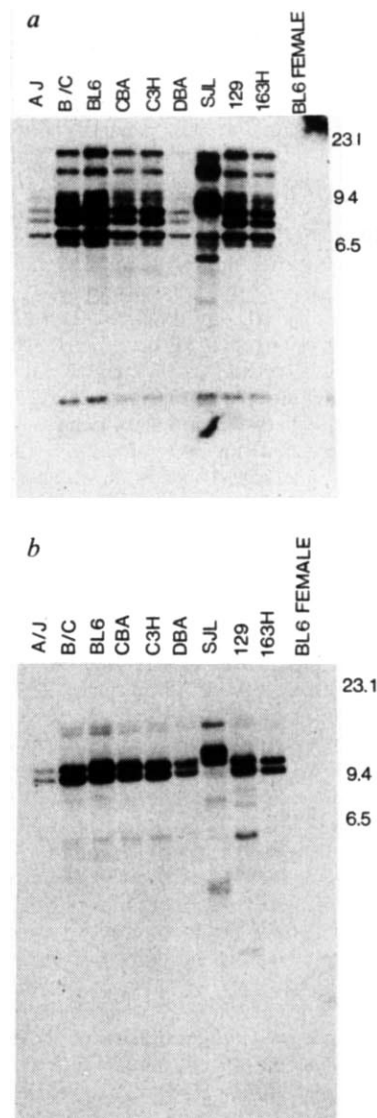


Fig. 3 Hybridization of pY353/B to genomic blots of nine established inbred laboratory mouse strains. Genomic blots were performed as described in Fig. 1 legend. The laboratory mouse strains are all maintained at the Institut Pasteur. Strains are (from left to right) A/J, BALB/c, C57BL/6, CBA/H, C3H, DBA/2, SJL, 129/sv, 163H and C57BL/6 female. Strain 163H is included as male fibroblasts from this line were used to flow-sort the Y (manuscript in preparation). a, The result of probing an *MspI* genomic blot (0.8% agarose) with pY353/B. All strains tested show the characteristic *M.m. musculus* banding pattern except SJL, which shows the *M.m. domesticus* pattern. The result is confirmed by b (*HindIII* digest, 0.7% agarose) in which all strains show the typical *M.m. musculus* pattern except SJL, which shows the characteristic *M.m. domesticus* pattern of hybridization. Re-probing the blots with a cloned mouse *H-2* cDNA probe showed that the lanes contained equivalent amounts of DNA. Therefore, the weaker hybridization intensity seen in the A/J and DBA/2 tracts is not due to loading differences but probably represents variations in copy number.

whereas pY353/B hybridized to all male DNAs. No differences could be detected between any of the mice on the *EcoRI* or *TaqI* blots but a clear dimorphism was seen on either the *MspI* or *HindIII* digests (Fig. 2a,b). Neither *TaqI* nor *HindIII* cuts within the probe whereas *MspI* does. The polymorphic forms coincided exactly with the semispecies designation. All *M.m. domesticus* mice carried one form, *M.m. musculus* the alternative form. There was no significant variation within the two groups using either enzyme. Thus, this probe can be used to discriminate

between the Y chromosomes of *M.m. domesticus* and *M.m. musculus*. Surprisingly, the Y chromosome carried by the laboratory reference strain C57BL/6 showed the characteristic patterns of *M.m. musculus* on both the *MspI* and *HindIII* digests. This finding was unexpected as it is generally assumed (and mitochondrial data support this assumption) that the laboratory strains were derived from *M.m. domesticus*.

To determine whether C57BL/6 is exceptional in this respect, we probed both *HindIII* and *MspI* genomic blots of nine commonly used established inbred laboratory mouse strains with pY353/B. The polymorphic patterns obtained (Fig. 3a,b) show that the Y carried by SJL is of the *M.m. domesticus* type whereas all the other strains tested (A/J, BALB/c, C57BL/6, CBA/HeJ, C3H, DBA/2, 129/sv and 163H) are clearly *M.m. musculus*-derived. The latter strains of mice must therefore be considered as genetic hybrids between two semispecies, with the maternal contribution derived from *M.m. domesticus* and the paternal genetic information derived from *M.m. musculus*. Further genetic studies (which are in progress) should reveal the extent of the hybrid characteristics and show whether they are restricted to the Y chromosome or not. These data suggest that the Y-chromosome polymorphism involving SJL reported recently¹² may result from the semispecies origin of the Y and not strict intraspecies polymorphism. In fact, no Y-chromosome RFLPs were detected in the new wild-derived *M.m. domesticus* and *M.m. musculus* lines; this could be related to the nature of the probe as it has been shown that the highly repeated cloned Bkm probe¹³ is extensively polymorphic both in wild and laboratory mouse strains (K. Jones, personal communication).

We consistently found that the intensity of hybridization of pY353/B to the A/J and DBA/2 inbred strains was 2–4 times less than that to the other strains, despite the blots containing approximately equal amounts of DNA from each mouse. As the hybridizations were performed under stringent washing conditions ($0.1 \times \text{SSC}$, 68°C), this difference probably represents a copy-number polymorphism rather than sequence divergence. Preliminary data using DNA from a mouse carrying a partially deleted Y chromosome (supplied by P. Burgoyne and E. Simpson) suggest that large blocks of DNA carrying this sequence are dispersed on the mouse Y chromosome (C.E.B., unpublished observations). It remains to be determined whether these copy-number polymorphisms arose during the establishment of the strains or whether they were present in the males which started the lines.

Historically, most of the laboratory mouse strains studied here (with the notable exception of SJL) are known to have originated from domesticated mice of North European and North American pet dealers and mouse fanciers. These mice were a mixture of locally caught North European (presumed *M.m. domesticus* origin) mice, subsequently carried across the Atlantic to North America, and imported 'fancy' Chinese and Japanese mice^{14,15}. If the *M.m. musculus* type Y chromosome described here originated from the European mice it would imply that some *M.m. domesticus* wild mice carry the *M.m. musculus* Y chromosome. Although we have been unable to show any such exchanges in the wild mice studied so far, they mainly originate from South-East Europe where such exchanges might not occur. It is known that *M.m. domesticus* and *M.m. musculus* mice hybridize in Europe (Fig. 1) and that in Scandinavia and Denmark there is a considerable one-way mitochondrial gene flow northeastwards from *M.m. domesticus* into *M.m. musculus*². In the southeastern part of the hybrid zone this flow is reversed, occurring in the *M.m. musculus* to *M.m. domesticus* direction⁴. We are at present surveying a very large European mouse population to determine whether there is a Y-chromosome gene flow between the semispecies.

Alternatively, the *M.m. musculus* type Y could have been introduced through the Chinese or Japanese mice. In this regard, Bonhomme *et al.* have recently shown, using numerous biochemical markers, that the Asiatic subspecies *Mus musculus molossinus* (found in China and Japan) is closely related to *M.m. musculus*⁵. Mitochondrial DNA polymorphisms also sup-

port this view^{6,16}. If this similarity extended to the Y chromosome, it could provide a simple explanation for the presence of a *M.m. musculus* type Y in the laboratory mouse strains, that is, being derived from the subspecies *M.m. molossinus*.

Whatever the origins of the Y chromosome, these data are of particular importance in studies involving hybrid genomes. Eicher *et al.* have reported that C57BL/6 differs from *M.m. domesticus* at an autosomal locus involved in gonadal determination and development (*tda-1*) as well as at the Y-localized testis-determining locus (*Tdy*)¹⁷. On the basis of the findings reported here, it seems probable that both *Tdy* and *tda-1* in C57BL/6 have co-evolved from *M.m. musculus*. The partial or complete sex reversal reported can be interpreted as a result of the inability of a *M.m. domesticus* Y chromosome to interact with the *M.m. musculus tda-1* locus. If this were true it would predict that the introduction of a Y from *M.m. musculus* into the C57BL/6 genome would not result in sex reversal, but the introduction of the SJL Y into the C57BL/6 genome would. Experiments are in progress to test this.

C.E.B. was in receipt of an INSERM Poste Orange grant. We thank Dr M. Goldberg and P. Metezeau for flow-sorting expertise, and acknowledge the support and advice of Drs M. Fellous, P. Avner, J. L. Guenet, P. Baldacci, G. Langsley, I. Garner and C. Rijnders.

Received 31 December 1984; accepted 26 February 1985.

- Palmer, J. D., Shields, C. R., Cohen, D. B. & Orton, T. J. *theor. appl. Genet.* **65**, 181–187 (1983).
- Ferris, S. D. *et al. Proc. natn. Acad. Sci. U.S.A.* **80**, 2290–2294 (1983).
- Powell, J. R. *et al. Proc. natn. Acad. Sci. U.S.A.* **80**, 492–495 (1983).
- Boursot, P. *et al. C.r. hebd. Séanc. Acad. Sci., Paris* **299**, 365–370 (1984).
- Bonhomme, F. *et al. Biochem. Genet.* **22**, 275–303 (1984).
- Ferris, S. D., Sage, R. D. & Wilson, A. C. *Nature* **295**, 163–165 (1982).
- Burgoyne, P. S. *Hum. Genet.* **61**, 85–90 (1982).
- Bishop, C. E. *et al. Nature* **303**, 831–832 (1983).
- Bishop, C. E. *et al. J. molec. Biol.* **173**, 403–417 (1984).
- Gueulaen, G. *et al. Nature* **307**, 172–173 (1984).
- Baron, B. *et al. Expl Cell Res.* **152**, 220–230 (1984).
- Lamar, E. E. & Palmer, E. *Cell* **37**, 171–177 (1984).
- Singh, L., Phillips, C. & Jones, K. W. *Cell* **36**, 111–120 (1984).
- Altman, P. L. & Katz, D. D. (eds) *Inbred and Genetically Defined Strains of Laboratory Animals* Pt 1 (Fedn Am. Soc. exp. Biol., Bethesda, Maryland, 1979).
- Klein, J. *Biology of the Mouse Histocompatibility-2 Complex* (Springer, New York, 1975).
- Yonekawa, H. *et al. Differentiation* **22**, 222–226 (1982).
- Eicher, E. M. & Washburn, L. *J. exp. Zool.* **228**, 297–304 (1983).

Simian virus 40 enhancer increases number of RNA polymerase II molecules on linked DNA

Richard Treisman* & Tom Maniatis

Department of Biochemistry and Molecular Biology, Harvard University, 7 Divinity Avenue, Cambridge, Massachusetts 02138, USA

Since the discovery of transcription enhancer sequences in the genomes of the DNA viruses simian virus 40 (SV40)^{1–3} and polyoma virus⁴, these elements have been shown to play an important part in the control of both viral and cellular gene expression^{5–10}. Enhancer elements act in *cis* to increase the amount of RNA produced from linked genes in a manner largely independent of distance and orientation^{1–3,11,12}. The mechanisms by which enhancers act are not understood; in particular, it is not known whether the enhancer-dependent increase in the level of stable RNA reflects an increase in the rate of transcription. To address this question, we have used an *in vitro* nuclear transcription assay^{13,14} to examine the effect of the SV40 enhancer on transcription of cloned human β -globin genes transiently introduced into HeLa cells¹⁵. We show here that the SV40 enhancer acts at least in part to increase the number of RNA polymerase II molecules transcribing the linked gene.

* Present address: MRC Laboratory for Molecular Biology, Hills Road, Cambridge CB2 2QH, UK.

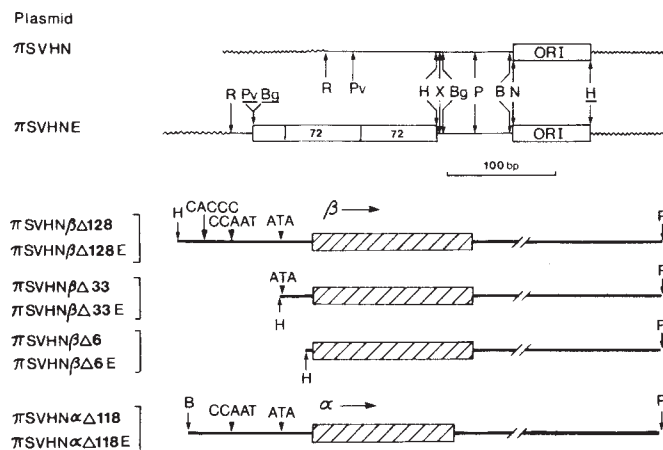


Fig. 1 Maps of the α - and β -globin gene plasmids used in HeLa cell transfection studies. Plasmids were constructed by standard procedures using the π VX miniplasmid as the parental plasmid^{16,25}. The two miniplasmids π SVHN (no enhancer) and π SVHNE (with enhancer) are shown at the top of the figure. Open boxes represent SV40 sequences, with the origin of DNA replication (ORI) and the enhancer sequence (72-bp repeat) indicated. The enhancer fragment contains SV40 nucleotides 272 (*Pvu*II site converted by *Bgl*III linker) to 100 (*Xho*I linker converted to *Hind*III). The π VX plasmid vector sequences are shown as wavy lines, the polylinker as thin lines. The human β - and α -globin genes used are shown at the bottom, with the exons indicated by the shaded boxes, and introns and flanking sequences by thick lines. The β -globin promoter elements CACCC, CCAAT and ATA^{18,19} are indicated. The β -globin plasmids lacking the SV40 enhancer, π SVHN $\beta\Delta 128$, π SVHN $\beta\Delta 33$ and π SVHN $\beta\Delta 6$, and the enhancer-containing β -globin plasmids π SVHN $\beta\Delta 128$ E, π SVHN $\beta\Delta 33$ E and π SVHN $\beta\Delta 6$ E contain the β -globin genes on a *Hind*III–*Pst*I fragment inserted between the corresponding sites in the polylinker sequences of π SVHN and π SVHNE. The β -globin deletion mutants were constructed as described previously^{22,26}, and contain 128, 33 and 6 bp of 5' flanking sequence of the β -globin gene. The α -globin plasmids π SVHN $\alpha\Delta 118$ and π SVHN $\alpha\Delta 118$ E were constructed by inserting the *Bam*HI (artificially introduced at position –118) to *Pst*I fragment of the human α_1 -globin gene between the *Bgl*III and *Pst*I sites of the two miniplasmids. The locations of relevant restriction enzyme cleavage sites are shown: R, *Eco*RI; Pv, *Pvu*II; H, *Hind*III; X, *Xba*I; Bg, *Bgl*III; P, *Pst*I; B, *Bam*HI; N, *Nco*I; sites underlined were destroyed during plasmid construction.

We first evaluated the effect of the SV40 enhancer on the accumulation of β -globin RNA using the S_1 nuclease mapping procedure. HeLa cells were transfected with plasmids carrying a human β -globin gene^{15,16} with or without the SV40 enhancer (Fig. 1), and stable RNA was analysed 48 h after transfection. The DNA fragment containing the enhancer does not include sequences sufficient for SV40 early or late promoter function. As an internal reference, a plasmid containing the human α -globin gene was also included in the transfection mix. Figure 2a shows that the presence of the SV40 enhancer next to the intact β -globin promoter increases the amount of correctly initiated RNA produced by two to three orders of magnitude^{1,16} (compare lane 1 with lanes 2–4).

We next isolated nuclei from transfected cells and incubated them with ³²P-labelled ribonucleoside triphosphates under conditions in which only limited elongation of previously initiated RNA chains occurs¹³ (see Fig. 3 legend). The amount of *in vitro*-synthesized, labelled RNA specific for the α - or β -globin genes or vector DNA was determined by hybridization of partially fragmented RNA to a large excess of double-stranded DNA fragments immobilized on nitrocellulose^{14,17}. The amount of radioactive RNA that hybridizes to a given fragment reflects the number of previously initiated RNA polymerase molecules present on the corresponding region of the transfected plasmid

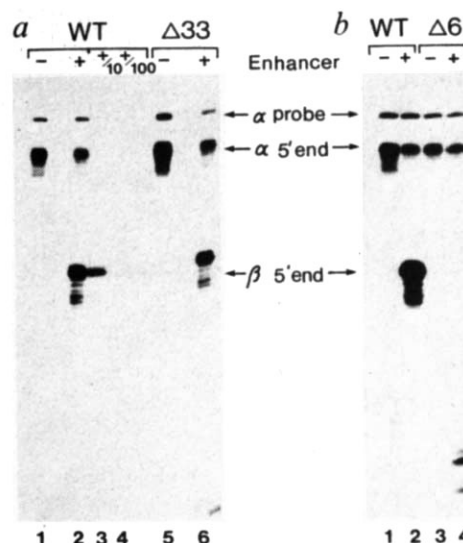


Fig. 2 Effect of the SV40 enhancer on the amount of accurately transcribed β -globin RNA in transfected HeLa cells. HeLa cells were transfected with 20 μ g of plasmid DNA per 9-cm dish and the RNA was analysed by S_1 nuclease mapping procedures²⁷. Each transfection included an equal mass of a human α -globin gene plasmid to serve as an internal reference. Unless otherwise stated, 40 μ g of total cell RNA was used for each assay. The positions of nuclease-resistant DNA fragments mapping the 5' ends of correctly initiated globin transcripts are indicated; – or + indicates the absence or presence, respectively, of the SV40 enhancer on the β -globin plasmid. The plasmids used contain β -globin genes with 128 (WT), 33 ($\Delta 33$) and 6 ($\Delta 6$) bp of 5' flanking sequence. **a**, Effect of the SV40 enhancer on the $\beta\Delta 128$ (intact promoter) and $\Delta 33$ (deleted upstream element) β -globin genes. The α -globin plasmid π SVHSA1 (ref. 15), which does not contain an SV40 enhancer, was used as a reference. Lane 1, plasmid π SVHN $\beta\Delta 128$; lane 2, plasmid π SVHN $\beta\Delta 128$ E; lane 3, plasmid π SVHN $\beta\Delta 128$ E, but the hybridization included only 4 μ g of RNA; lane 4, 1/10 of the S_1 nuclease-resistant counts loaded in lane 3; lane 5, plasmid π SVHN $\beta\Delta 33$; lane 6, plasmid π SVHN $\beta\Delta 33$ E. In lanes 3 and 5, the α -globin probe was omitted from the hybridization reaction. **b**, Effect of the SV40 enhancer on the $\beta\Delta 128$ (intact promoter) and $\beta\Delta 6$ (no promoter) globin genes. The α -globin gene plasmid π SVHPa2 (ref. 26), which contains an SV40 enhancer, was used as an internal reference. Lane 1, plasmid π SVHN $\beta\Delta 128$; lane 2, plasmid π SVHN $\beta\Delta 128$ E; lane 3, plasmid π SVHN $\beta\Delta 6$; lane 4, plasmid π SVHN $\beta\Delta 6$ E. RNA which apparently initiates within the 5'-noncoding region of the gene is indicated by the arrows to the right of lane 4. Cell culture, DNA transfection and S_1 nuclease mapping procedures have been described previously^{15,28}. The single-stranded [5'-³²P]-labelled probes were: β -globin, *Hae*III–*Mst*II*, nucleotides –76 to +70; α -globin, *Hha*I (artificially constructed; R.T., unpublished) to *Bss*HII*, nucleotides –11 to +97.

DNA¹⁴. The results of such a nuclear transcription experiment are shown in Fig. 3; in this experiment HeLa cells were transfected with β -globin plasmids that differ by the presence or absence of the SV40 enhancer sequence, together with a plasmid containing the human α -globin gene as an internal reference. Both globin and vector-specific RNAs were readily detectable after *in vitro* elongation of preinitiated RNA, and a substantial increase in the amount of β -globin specific RNA was observed when the gene was linked to the enhancer (Fig. 3a, compare lanes 1 and 2). A similar effect was observed with the α -globin gene using the β -globin plasmid as an internal reference (Fig. 3a, compare lanes 2 and 3). Examination of stable RNA showed that in the absence of a linked enhancer, the α -globin gene produced readily detectable levels of correctly initiated RNA, and that the presence of the enhancer at the 5' side of the gene resulted in a 30–50-fold increase in the amount of RNA produced (data not shown). When 2 μ g ml^{–1} α -amanitin was included in the nuclear transcription reaction, no globin- or vector-specific transcripts were synthesized (Fig. 3b, compare

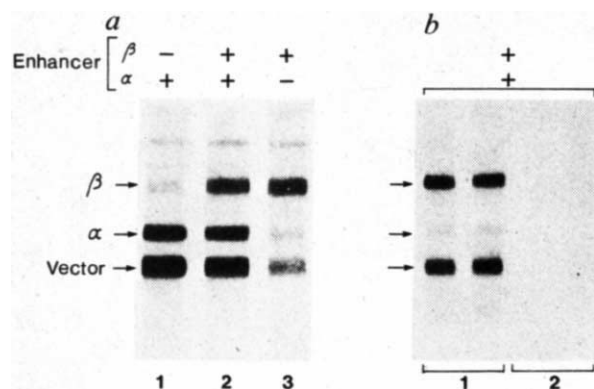


Fig. 3 Effect of the SV40 enhancer on nuclear transcription of transfected α - and β -globin plasmids. Each 'lane' is an autoradiogram of a nitrocellulose strip carrying fragments of β -globin, α -globin or vector DNA, after hybridization to RNA synthesized *in vitro*. The identities of the fragments and the presence (+) or absence (-) of the SV40 enhancer on each of the transfected plasmids are indicated. *a*, Effect of the SV40 enhancer on α - and β -globin nuclear transcription. HeLa cells were transfected with the following plasmids: lane 1, π SVHN $\beta\Delta 128$ + π SVHN $\alpha\Delta 118$ E; lane 2, π SVHN $\beta\Delta 128$ E + π SVHN $\alpha\Delta 118$ E; lane 3, π SVHN $\beta\Delta 128$ E + π SVHN $\alpha\Delta 118$. *b*, Effect of α -amanitin on nuclear transcription of transfected plasmids. HeLa cells were transfected with π SVHN $\beta\Delta 128$ E and π SVHP $\alpha 2$ (ref. 26), and nuclear transcription reactions performed either in the absence (lane 1) or presence (lane 2) of $2 \mu\text{g ml}^{-1}$ α -amanitin. Duplicate strips are shown.

Methods. Nuclei were isolated by hypotonic lysis using a Dounce homogenizer (type A pestle) as described previously¹³. Cell lysis was evaluated by Trypan blue staining. Nuclei from two 9-cm plates of transfected cells were used for each reaction. Isolated nuclei were resuspended at 0°C in a total volume of $100 \mu\text{l}$, comprising $50 \mu\text{l}$ $2\times$ transcription buffer¹³ (50 mM dithiothreitol, 180 mM KCl, 10 mM MgCl_2 , 20 mM Tris-HCl, 50% glycerol, pH 7.8), brought to 0.25 mM ATP, 0.125 mM GTP, 0.125 mM CTP, 100 μCi [α - ^{32}P]UTP (410 Ci mmol^{-1} ; Amersham) was added to each reaction. The nuclei were incubated at 37°C for 5 min; no incorporation occurred after this time. Under these conditions, pre-initiated RNA chains are elongated by 300–500 nucleotides¹³. RNA was isolated by protease K treatment, followed by two extractions with phenol/chloroform as described previously²⁷; residual DNA was removed by treatment with RNase-free DNase. Approximately 5×10^6 to 10^7 c.p.m. RNA were routinely obtained. Fragmentation of RNA by partial alkali hydrolysis was as described elsewhere²⁹. For analysis of the *in vitro*-synthesized RNA by strip blot hybridization¹⁴, plasmids π SVHN $\beta\Delta 128$ and π SVHN $\alpha\Delta 118$ (25 μg each) were digested with *Xba*I and *Pst*I and the fragments resolved on a 10 cm-wide 1.2% agarose gel, then transferred to nitrocellulose³⁰. Duplicate 1 cm-wide strips of nitrocellulose were then hybridized to equal numbers of counts as described⁴⁰, washed, and exposed to Kodak X-AR5 film.

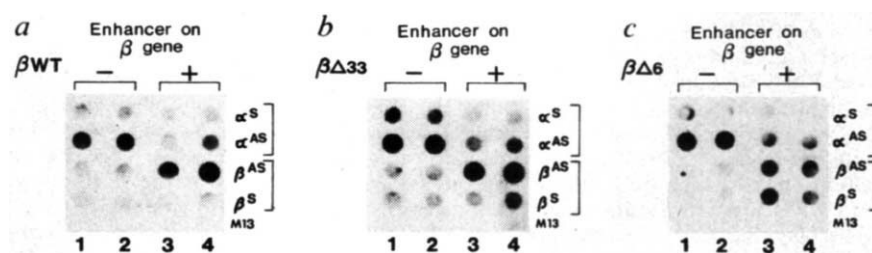
lanes 1 and 2), indicating that the *in vitro* transcription is due to RNA polymerase II. We also verified that reinitiation does not occur in the system by performing nuclear incubations in the presence of 0.5% sarkosyl or 1 mg ml^{-1} heparin. Although these additions slightly diminished the total *in vitro* RNA synthesis, little effect on *in vitro* synthesis of α - and β -globin specific RNA was observed (data not shown).

These results demonstrate that the SV40 enhancer causes a substantial increase in the number of RNA polymerase II molecules present on linked DNA. However, the experiment does not distinguish between *in vitro* elongation of transcripts initiated at the β -globin promoter and the elongation of nascent transcripts initiated at other sites in the plasmid DNA. To address this problem, we examined the effect of promoter deletions on β -globin gene specific transcription in isolated nuclei. Plasmids were constructed in which the enhancer is located at the 5' side of truncated β -globin genes containing either 33 base pairs (bp) of 5' flanking sequence (which includes the ATA box, but lacks the upstream promoter elements CACCC and CCAAT^{18,19}), (plasmids π SVHN $\beta\Delta 33$ and π SVHN $\beta\Delta 33$ E; Fig. 1) or only 6 bp of 5' flanking sequence (which lacks all promoter elements), (plasmids π SVHN $\beta\Delta 6$ and π SVHN $\beta\Delta 6$ E; Fig. 1). The transcription of these plasmids in transfected HeLa cells was then analysed by S_1 nuclease mapping and by the nuclear transcription assay. The S_1 analysis (Fig. 2) shows that (1) neither of the truncated β -globin genes directs the synthesis of detectable levels of correctly initiated β -globin RNA in the absence of a linked SV40 enhancer sequence (Fig. 2a, lane 5; Fig. 2b, lane 1); (2) deletion of sequences 5' to the ATA box has essentially no effect on the level of correctly initiated RNA produced when the enhancer is in close proximity to the ATA box, indicating that the enhancer can compensate for the deletion of upstream promoter sequences as reported previously for other promoters^{20,21} (Fig. 2a, lane 6). (We note, however, that when the SV40 enhancer is located farther upstream from the $\beta\Delta 33$ gene, or on its 3' side, virtually no transcription is observed²².) (3) Deletion of all promoter elements results in the loss of correctly initiated β -globin RNA and the concomitant appearance of a small amount of RNA that is apparently initiated at, or spliced to, sites in the 5'-untranslated region of the gene (Fig. 2b, lane 4).

Figure 4 shows a nuclear transcription analysis of the β -globin promoter deletions. The RNA synthesized *in vitro* was hybridized to the separate strands of the β - and α -globin genes cloned into M13mp8. The linkage of the SV40 enhancer to the intact β -globin gene increases specifically the amount of transcription of the coding (or 'anti-sense') DNA strand (Fig. 4a, compare strips 1, 2 with 3, 4). A small amount of transcription of the noncoding strands of both β - and α -globin genes was detected, even in the absence of a linked SV40 enhancer. We also observed a decrease in the number of RNA polymerase molecules on the α -globin reference gene when the test β -globin gene was linked

Fig. 4 Effect of the SV40 enhancer on nuclear transcription of β -globin promoter deletion mutants. Each panel contains pairs of duplicate strips of dot blots on which are spotted recombinant M13 phage carrying the sense (S) and anti-sense (AS) strands of the α - and β -globin genes. Bona fide transcripts hybridize with the anti-sense strand. HeLa cells were transfected with π SVHP $\alpha 2$ (ref. 26) or *a*, plasmid π SVHN $\beta\Delta 128$ (strips 1, 2) or π SVHN $\beta\Delta 128$ E (strips 3, 4); *b*, π SVHN $\beta\Delta 33$ (strips 1, 2) or π SVHN $\beta\Delta 33$ E (strips 3, 4); *c*, π SVHN $\beta\Delta 6$ (strips 1, 2) or π SVHN $\beta\Delta 6$ E (strips 3, 4).

Methods. Nuclear transcriptions, RNA preparation and hybridization conditions were as described in Fig. 3 legend. Recombinant M13 phage were constructed by insertion of the 1.5-kb *Pst*I fragment of the human α_1 -globin gene, or a 2.2-kb *Pst*I fragment of the human β -globin gene (from a *Pst*I linker at the endpoint of the -33 deletion to the *Pst*I site 500 bp on the 3' side of the gene) into M13mp8 in either orientation, by standard procedures. The M13 DNAs (5 μg) were spotted onto nitrocellulose in $10\times$ SSC with a 'Manifold' apparatus (Schleicher & Schuell).



to the enhancer, suggesting that co-transfected enhancers can compete for limiting factors in the *in vitro* transcription assay. We do not observe this effect at the level of stable RNA under our conditions, indicating that factors that may be limiting *in vitro* are not limiting *in vivo*. Others have reported enhancer competition *in vivo* using different transcription units and transfection conditions²³. In agreement with the observation that deletion of the β -globin upstream elements in plasmid π SVHN $\beta\Delta$ 33E does not affect the accumulation of stable RNA, we find that the behaviour of this plasmid in the nuclear transcription assay is identical to that observed with the plasmid carrying the intact β -globin gene (Fig. 4, compare *a* and *b*). A strikingly different result was obtained when all the β -globin promoter elements were removed (Fig. 4c). Again, a small amount of *in vitro* transcription of both strands of the truncated gene was detected even in the absence of a linked enhancer sequence. However, linkage of the SV40 enhancer sequence to the gene lacking promoter elements resulted in a small increase in transcription of both DNA strands (Fig. 4c, compare strips 1, 2 with 3, 4); this increase may result from the activation of normally silent 'cryptic' start sites such as that observed in the 5'-untranslated region of the β -globin gene (Fig. 2*b*, lane 4). We conclude that enhancer effects seen in the nuclear transcription assays do depend on known promoter elements of the β -globin gene, and therefore relate to the production of bona fide β -globin RNA.

We have demonstrated here that the SV40 enhancer causes an increase in the number of RNA polymerase II molecules engaged in the transcription of linked DNA. The increase is dependent on known components of the β -globin gene promoter^{18,19}. It was reported previously that deletion of the adenovirus 2 E1A enhancer causes a reduction in synthesis of E1A RNA measured in pulse labelling experiments²⁴. Our observations exclude the possibility that the enhancer acts solely to allow stabilization of nascent transcripts. The magnitude of the enhancer effect observed in the nuclear transcription assay appears to be smaller than the two to three orders of magnitude effect on the level of stable RNA (compare Figs 2 and 3). Moreover, the 10–20-fold difference in the effect of the SV40 enhancer on the production of stable α - and β -globin RNA¹⁵ is apparently not observed in the nuclear transcription assay. However, accurate quantitation of enhancer effects in the *in vitro* nuclear transcription assay is difficult. At least some of the low level of α -globin nuclear transcription observed in the absence of a linked enhancer sequence (Fig. 3*a*) must be due to elongation of correctly initiated RNA, as this plasmid does produce easily detectable levels of stable α -globin RNA (data not shown; see ref. 15). It is possible, however, that the low level of β -globin nuclear transcription observed in the absence of a linked enhancer does not represent elongation of RNA initiated at the correct β -globin cap site. This possibility is suggested by the observation that nuclear transcription in the absence of the enhancer is not significantly dependent on the β -globin promoter (Fig. 4, compare *a* and *c*). The magnitude of the effect we observe therefore provides a minimum estimate of the actual effect of the enhancer on polymerase loading. We cannot exclude the possibility that the enhancer also acts at levels other than transcription initiation to increase the production of stable RNA.

R.T. was supported by a special fellowship from the Leukemia Society of America Inc. This work was supported by a NIH grant to T.M.

Received 8 February; accepted 7 March 1985.

1. Banerji, J., Rusconi, S. & Schaffner, W. *Cell* **27**, 299–308 (1981).
2. Moreau, P. *et al. Nucleic Acids Res.* **9**, 6047–6068 (1981).
3. Fromm, M. & Berg, P. *J. molec. appl. Genet.* **1**, 457–481 (1982).
4. de Villiers, J., Olson, L., Tyndall, C. & Schaffner, W. *Nucleic Acids Res.* **10**, 7965–7976 (1982).
5. Gruss, P. & Khoury, G. *Cell* **33**, 313–314 (1983).
6. Banerji, J., Olson, L. & Schaffner, W. *Cell* **33**, 729–740 (1983).
7. Gillies, S. D., Morrison, S. L., Oi, V. & Tonegawa, S. *Cell* **33**, 717–728 (1983).
8. Neuberger, M. S. *EMBO J.* **1**, 1373–1378 (1983).
9. Chandler, V. L., Maler, B. A. & Yamamoto, K. R. *Cell* **33**, 489–499 (1983).
10. Gluzman, Y. & Shenk, T. *Enhancers and Eukaryotic Gene Expression* (Cold Spring Harbor Laboratory, New York, 1983).

11. Wasylyk, B., Wasylyk, C., Augerau, P. & Chambon, P. *Cell* **32**, 503–514 (1983).
12. de Villiers, J., Olson, L., Banerji, J. & Schaffner, W. *Cold Spring Harb. Symp. quant. Biol.* **47**, 911–919 (1982).
13. Weber, J., Jelinek, W. & Darnell, J. E. *Cell* **10**, 611–616 (1977).
14. Groudine, M., Peretz, M. & Weintraub, H. *Molec. Cell Biol.* **1**, 281–288 (1981).
15. Treisman, R. H., Green, M. R. & Maniatis, T. *Proc. natn. Acad. Sci. U.S.A.* **80**, 7428–7432 (1983).
16. Little, P. F. R., Treisman, R., Beirut, L., Seed, B. & Maniatis, T. *Molec. Biol. Med.* **1**, 473–488 (1984).
17. Birg, F., Favaloro, J. & Kamen, R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 3128–3132 (1977).
18. Dierks, P. *et al. Cell* **32**, 695–706 (1983).
19. Grosveld, G. C., de Boer, E., Shewmaker, C. K. & Flavell, R. A. *Nature* **295**, 120–126 (1982).
20. Pelham, H. R. B. *Cell* **30**, 517–528 (1982).
21. Hen, R., Sassone-Corsi, P., Corden, J., Gaub, M. P. & Chambon, P. *Proc. natn. Acad. Sci. U.S.A.* **79**, 7132–7136 (1982).
22. Green, M. R., Treisman, R. H. & Maniatis, T. *Cell* **35**, 137–148 (1983).
23. Scholer, H. R. & Gruss, P. *Cell* **36**, 403–411 (1984).
24. Hearing, P. & Shenk, T. *Cell* **30**, 517–528 (1983).
25. Seed, B. *Nucleic Acids Res.* **11**, 2427–2443 (1983).
26. Treisman, R. H., Orkin, S. H. & Maniatis, T. *Nature* **302**, 591–596 (1983).
27. Favaloro, J. M., Treisman, R. H. & Kamen, R. *Meth. Enzym.* **65**, 718–749 (1980).
28. Busslinger, M., Moschonas, N. & Flavell, R. A. *Cell* **27**, 289–298 (1981).
29. Hofer, E. & Darnell, J. E. *Cell* **23**, 585–593 (1981).
30. Southern, E. M. *J. molec. Biol.* **98**, 503–517 (1975).

Simian virus 40 enhancer increases RNA polymerase density within the linked gene

Frank Weber & Walter Schaffner

Institut für Molekularbiologie II der Universität Zürich, Hönggerberg, CH-8093 Zürich, Switzerland

Enhancers are regulatory DNA elements, usually about 200 base pairs (bp) long, which are able to stimulate transcription of linked genes in eukaryotic cells^{1,2}. This activation can be exerted over large distances, and from a position 5' or 3' to the gene¹. Enhancers have been identified in viral genomes^{1–9} and cellular genes^{10–15}. Using a transient expression assay, we have analysed transcription of the rabbit β -globin gene and the thymidine kinase gene from herpes simplex virus with and without a simian virus 40 (SV40) enhancer. S₁ nuclease mapping shows a high level of specific transcripts when the genes are linked to the enhancer. To determine whether this increased number of transcripts is due to a higher transcription rate, or perhaps to a shift from nonspecific to specific initiation, we have performed run-on transcription assays with isolated nuclei. Our results, presented here, demonstrate that the SV40 enhancer increases the RNA polymerase density within the linked gene. Therefore, enhancers apparently increase the rate of transcription initiation without influencing the specificity of initiation.

To study the enhancer effect on polymerase II transcription units, we constructed clones containing the rabbit β -globin gene and the herpes simplex virus (HSV) thymidine kinase (*tk*) gene with and without the SV40 enhancer (see Fig. 1).

The rabbit β -globin and the HSV *tk* genes, which were previously shown to respond to an enhancer^{1,8}, were cloned as 2-kilobase (kb) fragments into the plasmid pUC9 (ref. 16). The SV40 72-bp repeat enhancer was cloned as a 195-bp fragment^{17,18} ~1.5 kb downstream of the respective promoters.

To demonstrate enhancer-dependent expression of the two genes, we transfected the clones into human HeLa cells, using the calcium phosphate co-precipitation method^{18,19}. RNA was extracted after 2 days and analysed by S₁ nuclease mapping. The S₁ analysis clearly shows that the enhancer stimulates the level of transcription of the β -globin and *tk* genes ~40-fold and 15-fold, respectively (Fig. 2). The effect on β -globin transcription, however, is less pronounced than in previous experiments in which stimulation by the SV40 enhancer was ~200-fold¹. This difference results mainly from an increased level of transcripts in the absence of an enhancer when using the pUC vector instead of the previously used pBR322 (data not shown), and was not analysed further.

An analysis of both the 5' and 3' ends of the rabbit β -globin transcripts demonstrates correct initiation, 3' processing and

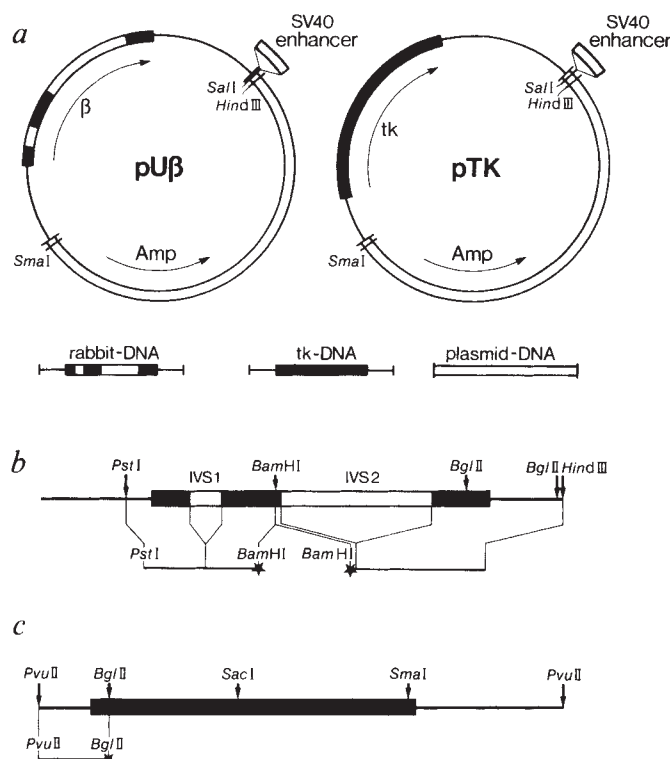


Fig. 1 Maps of recombinant DNAs and schematic representation of the S_1 nuclease mapping probes used. *a*, pU β : the 2-kb *Bgl*II fragment of the rabbit β_1 -globin gene²⁹ was cloned into the *Bam*HI site of pUC9 (ref. 16) so that the direction of transcription of the globin gene was opposite to that of the *lacZ* gene contained in the plasmid. pTK: the 2-kb *Pvu*II fragment of the HSV *tk* gene³⁰ was first subcloned into the polylinker of pUC8 and thereafter excised as a *Bam*HI fragment for insertion into pUC9. The clones containing a 195-bp insert of the SV40 enhancer^{17,18} were called pU β E and pTK-E, respectively. *b*, A β -globin gene lacking both intervening sequences was used to prepare the radioactive probes. To map β -globin transcripts, DNA was end-labelled at the *Bam*HI site using either T4 polynucleotide kinase (for 5'-specific probe) or the Klenow fragment of DNA polymerase I (3'-specific probe). *c*, DNA was labelled at the *Bgl*II site to obtain a 5'-specific *tk* gene probe.

splicing of the two introns during maturation of messenger RNA, independent of whether an enhancer was present or not (Fig. 2*a, b*). Also, the *tk* mRNA expressed from pTK and pTKE has a correct 5' end (Fig. 2*c*). The enhancer, therefore, does not apparently influence the specificity of initiation or the processing of the pre-mRNAs.

However, S_1 nuclease analysis measures only the steady-state level of stable transcripts. In particular, the presence of a clean S_1 -protected band does not rule out the possible existence of a wealth of nonspecific unstable transcripts from the same gene.

To determine whether an enhancer shifts the equilibrium from nonspecific to specific transcripts²⁰, or whether it increases the rate of gene transcription, we measured the RNA polymerase II density. To this end, we have determined the enhancer effect by a run-on transcription experiment in isolated nuclei^{21,22}; this technique is widely used to determine the polymerase density on an endogenous gene of defined copy number.

It is known that the enhancer effect is not due to differential uptake of the DNA¹; in particular, the effect is also seen when defined numbers of molecules are injected into cells^{2,8} and when template molecules with and without enhancer are replicated to high copy numbers using COS cells (refs 20, 23, 24 and T. Gerster and W.S., unpublished). Therefore, the run-on technique is also applicable to the study of transcription in a transient assay²⁵. The combination of the transient expression system with a nuclear run-on assay offers several advantages: for example,

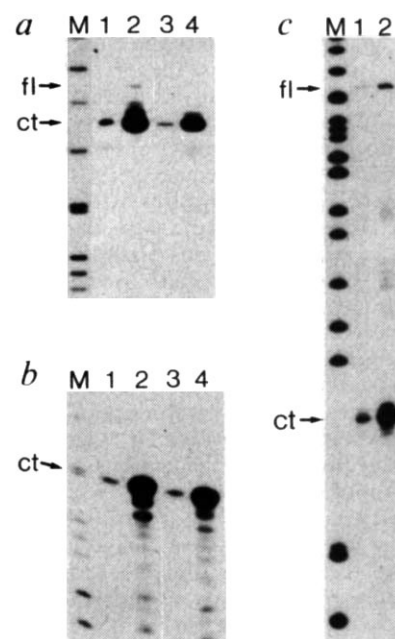


Fig. 2 Enhanced β -globin and thymidine kinase transcription. *a*, Hybridization of a 5'-specific β -globin probe (Fig. 1*b*) to 20 μ g of RNA from HeLa cells transfected with: recombinant pU β (lane 1); pU β E (lane 2); pU β , linearized with *Hind*III (lane 3); pU β E, linearized with *Hind*III (lane 4). M, End-labelled marker DNA fragments (pBR322 cleaved with *Hpa*II). fl, Full-length probe (453 nucleotides); ct, correct terminus (354 nucleotides). *b*, Hybridization of a 3'-specific β -globin probe to 20 μ g RNAs of the same transfections as in *a*. ct, 240 nucleotides. *c*, Hybridization of a 5'-specific *tk* gene probe (Fig. 1*c*) to 20 μ g RNA from HeLa cells transfected with recombinant pTK (lane 1) and pTKE (lane 2). fl, 270 nucleotides; ct, 54 nucleotides.

Methods. Tissue culture cell lines were grown and transfected as described previously¹⁸. As a control for transfection efficiency, a 'tracer' gene was added to one of two duplicate plates: (1) A clone containing the SV40 T-antigen gene under the control of the adenovirus major late promoter and an SV40 enhancer was mixed 1:100 with the appropriate β -globin clone. Two days post-transfection, the cells in the plate co-transfected with tracer were fixed and stained for T-antigen by immunofluorescence. (2) The cells in the plate without tracer were collected for RNA extraction. Similar numbers of stained nuclei were observed independent of the globin construct used (523 T-antigen-positive per 14,000 cells and 509 positive per 13,000 cells with pU β and pU β E, respectively). This finding was corroborated by Southern blot analysis of nuclear DNA from transfected HeLa cells, which yielded equal amounts of cloned DNA present within the nuclei (not shown). RNA analysis was performed as described elsewhere^{31,32}. End-labelled single-stranded DNA (see Fig. 1*b, c*) was hybridized to unlabelled RNA, treated with S_1 nuclease, denatured, fractionated by gel electrophoresis and autoradiographed.

the rate of transcription within different cloned and mutagenized genes can be compared, which is particularly useful with mutants possibly affecting mRNA stability or transcription termination of mRNA (for the latter case see below and Fig. 3).

Nuclei of HeLa cells transfected with our β -globin and *tk* clones were isolated and digested with RNase A. Pre-initiated RNA was then elongated *in vitro* for 10 min in the presence of [α -³²P]rGTP, giving rise to labelled RNA molecules ~200–300 nucleotides long (Fig. 3*a*). As RNA polymerase II appears to protect ~60 nucleotides from nuclease digestion, the *in vivo* elongation rate is ~15–20 nucleotides per min in HeLa nuclei; this rate is much lower than that found *in vivo* but is in good agreement with previous *in vitro* studies^{22,26}.

The labelled RNA was then hybridized to dot-blots of cloned DNA. The dot hybridization signals in the run-on assays are

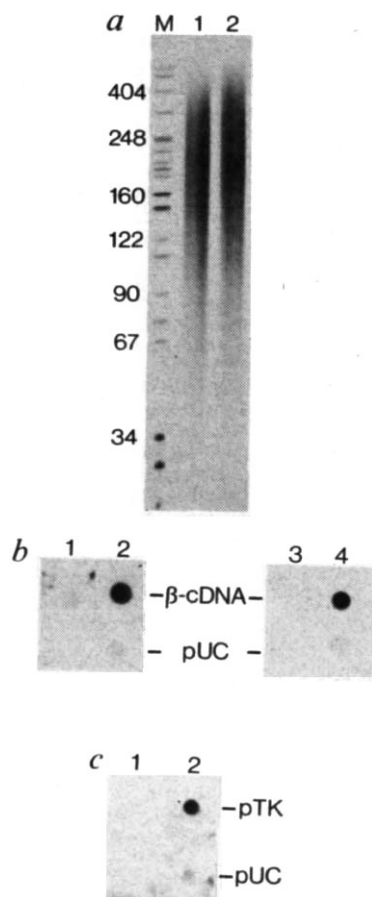


Fig. 3 Run-on assays also reveal the enhancer effect. *a*, RNA from nuclei which had been treated with RNase A and labelled *in vitro* was isolated and subjected to denaturing polyacrylamide (6%) gel electrophoresis. Cells transfected with pUβ (lane 1) and pUβE (lane 2) show a similar size distribution of *in vitro* elongated nuclear RNA; M, end-labelled marker DNA fragments (pBR322 cleaved with *Hpa*II). *b*, Unlabelled DNA of pUC9 and pcβ (rabbit β-globin cDNA) was denatured, spotted onto nitrocellulose membrane and hybridized to 4×10^6 c.p.m. (lanes 1 and 2) or 1.2×10^6 c.p.m. (lanes 3 and 4) of nuclear RNA from cells transfected with: pUβ (lane 1); pUβE (lane 2); pUβ linearized with *Hind*III (lane 3); pUβE linearized with *Hind*III (lane 4). *c*, Hybridization of pUC and pTK DNA to 6×10^6 c.p.m. of *in vitro* elongated nuclear RNA from cells transfected with pTK (lane 1) and pTKE (lane 2). HeLa cells were transfected as described previously¹⁸. The nuclei were isolated 36 h post-transfection as described²¹, except that divalent cations were omitted from the buffers. Run-on assays were performed as published elsewhere²².

expected to be proportional to the polymerase densities as RNase A treatment of the nuclei before *in vitro* elongation removes the pre-existing different amounts of unlabelled RNA and thus avoids competition of the mature mRNA with nascent transcripts²². We found that nuclear RNA from cells transfected with pUβE or pTKE hybridizes strongly to rabbit β-globin complementary DNA or *tk* DNA, respectively, whereas nuclear RNA from pUβ- or pTK-transfected cells hybridizes only weakly (Fig. 3b, c), in agreement with the *S*₁ nuclease mapping assays shown in Fig. 2.

The nuclear RNAs hybridize ~10% as strongly to plasmid vector DNA as β-globin or *tk* gene sequences (Fig. 3b, c), indicating (1) that the enhancer primarily stimulates transcription from the β-globin or *tk* promoter rather than from pseudopromoters within plasmid sequences, and (2) that there is efficient transcription termination preventing readthrough into plasmid DNA.

In agreement with our *S*₁ nuclease mapping data (Fig. 2), similar results were obtained whether transfection was per-

formed using circular or linear DNA (Fig. 3a, b; the latter is probably ligated within the transfected cells¹⁸).

Thus, our results indicate strongly that an enhancer increases the RNA polymerase II density on a gene to which it is linked, that is, the enhancer facilitates transcription by increasing the frequency of initiation. In particular, the enhancer does not seem to shift the equilibrium from nonspecific to specific transcripts, as could have been inferred from the work of Humphries *et al.*²⁰. These authors studied the expression of human α-, β- and δ-globin genes using vectors containing an SV40 replicon²⁷ and were struck by the high levels of nonspecific transcripts in the absence of an enhancer. However, we consider it likely that the high level of nonspecific β-globin transcripts is a peculiarity of their assay conditions. With replicating SV40 templates, there can be a considerable fraction of nonspecific transcripts from multiple scattered initiation points²⁴ which are perhaps related to the multiple initiation sites of SV40 late transcripts²⁸.

We thank Ueli Schibler (Epalinges) for technical advice, Fritz Ochsenbein for preparing the graphs, and Maria Jasin and Max L. Birnstiel for critical reading of the manuscript. This work was supported by the Kanton of Zürich and the Swiss National Research Foundation.

Received 24 December 1984; accepted 25 February 1985.

1. Banerji, J., Rusconi, S. & Schaffner, W. *Cell* **27**, 299-308 (1981).
2. Moreau, P. *et al. Nucleic Acids Res.* **9**, 6047-6068 (1981).
3. de Villiers, J. & Schaffner, W. *Nucleic Acids Res.* **9**, 6251-6264 (1981).
4. Lusky, M., Berg, L., Weiher, H. & Botchan, M. *Molec. cell. Biol.* **3**, 1108-1122 (1983).
5. Lang, J. C., Spandidos, D. A. & Wilkies, N. M. *EMBO J.* **3**, 389-395 (1984).
6. Boshart, M. *et al. Cell* (in the press).
7. Laimins, L. A., Khoury, G., Gorman, C., Howard, B. & Gruss, P. *Proc. natn. Acad. Sci. U.S.A.* **79**, 6453-6457 (1982).
8. Luciw, P. A., Bishop, J. M., Varus, H. E. & Capecci, M. R. *Cell* **33**, 705-716 (1983).
9. Hearing, P. & Shen, T. *Cell* **33**, 695-703 (1983).
10. Banerji, J., Olson, L. & Schaffner, W. *Cell* **33**, 729-740 (1983).
11. Gillies, S. D., Morrison, S. L., Oi, V. T. & Tonegawa, S. *Cell* **33**, 717-728 (1983).
12. Neuberger, M. S. *EMBO J.* **2**, 1373-1378 (1983).
13. Picard, D. & Schaffner, W. *Nature* **307**, 80-82 (1984).
14. Gillies, S. D., Folsom, V. & Tonegawa, S. *Nature* **310**, 594-597 (1984).
15. Walker, M. D., Edlund, T., Boulet, A. M. & Rutten, W. J. *Nature* **306**, 557-561 (1983).
16. Vieira, J. & Messing, J. *Gene* **19**, 259-268 (1982).
17. de Villiers, J., Olson, L., Banerji, J. & Schaffner, W. *Cold Spring Harb. Symp. quant. Biol.* **47**, 911-919 (1982).
18. Weber, F., deVilliers, J. & Schaffner, W. *Cell* **36**, 983-992 (1984).
19. Graham, F. L. & van der Eb, A. J. *Virology* **52**, 456-467 (1973).
20. Humphries, R. K., Ley, T., Turner, P., Moulton, A. D. & Nienhuis, A. W. *Cell* **30**, 173-183 (1982).
21. Groudine, M., Peretz, M. & Weintraub, H. *Molec. cell. Biol.* **1**, 281-288 (1981).
22. Schibler, U., Hagenbüchle, O., Wellauer, P. K. & Pittet, A. C. *Cell* **33**, 501-508 (1983).
23. Gluzman, Y. *Cell* **23**, 175-182 (1981).
24. Treisman, R., Green, M. R. & Maniatis, T. *Proc. natn. Acad. Sci. U.S.A.* **80**, 7428-7432 (1983).
25. Leff, T. *et al. Proc. natn. Acad. Sci. U.S.A.* **81**, 4381-4385 (1984).
26. Gocke, E., Leer, J. C., Nielsen, O. F. & Westergaard, O. *Nucleic Acids Res.* **5**, 3993-4006 (1978).
27. Mellon, P., Parker, V., Gluzman, Y. & Maniatis, T. *Cell* **27**, 279-288 (1981).
28. Contreras, R., Gheysen, D., Knowland, J., van de Voorde, A. & Fiers, W. *Nature* **300**, 500-505 (1982).
29. Maniatis, T. *et al. Cell* **15**, 687-701 (1978).
30. Wagner, M. J., Sharp, J. A. & Summers, W. C. *Proc. natn. Acad. Sci. U.S.A.* **78**, 1441-1445 (1981).
31. Weaver, R. F. & Weissmann, C. *Nucleic Acids Res.* **7**, 1175-1193 (1979).
32. de Villiers, J. & Schaffner, W. *Methods in Nucleic Acid Biochemistry* **B507**, 1-20 (Elsevier, Amsterdam, 1983).

Erratum

Functions of the canal system in the rotaliid foraminifer, *Heterostegina depressa*

R. Röttger, M. Spindler, R. Schmaljohann,
M. Richwien & M. Fladung
Nature **309**, 789-791 (1984)

ON page 789, the last sentence of the second paragraph should read "We have examined the symbiont-bearing Recent nummulitid, *H. depressa*, in which the canal system within the chamber walls and within the marginal cord replaces the primary aperture seen in other foraminifera."

Physics

Handbook of Fluids in Motion. N.P. CHEREMISINOFF and R. GUPTA (eds). *Ann Arbor: 1983. Pp. 1202. ISBN 0-250-40458-3. £75.*

Quark Matter Formation and Heavy Ion Collisions. Proceedings of the Bielefeld Workshop held May 1982. M. JACOB and H. SATZ (eds). *World Scientific: 1982. Pp. 586. Hbk ISBN 9971-950-46-4; pbk ISBN 9971-950-47-2. Np.*

Selected Lectures Hawaii Tropical Conference in Particle Physics, Vol. 1 and 2. S. PAKVASA and S.F. TUAN (eds). *World Scientific: 1982. Pp. 991. Vol. 1 Hbk ISBN 9971-950-36-7; pbk ISBN 9971-950-37-5; Vol. 2 Hbk ISBN 9971-950-16-2; pbk ISBN 9971-950-17-0. Vol. 1 Hbk £36, \$51; pbk £17.24; Vol. 2 Hbk £36, \$51; pbk £17, \$24.*

Chemistry

Chemical Nomenclature Usage. R. LEES and A.F. SMITH (eds). *Ellis Horwood/Halsted: 1983. Pp. 172. ISBN 0-85312-368-3/0-470-27447-6. Hbk £19.50, \$34.70; pbk np.*

Practice of Thin Layer Chromatography, 2nd Edn. J.C. TOUCHSTONE and M.F. BOBBINS. *Wiley: 1983. Pp. 405. ISBN 0-471-09766-7. £38, \$53.20.*

Technology

Efficiency of Manufacturing Systems. NATO Conference Series, Vol. 14. B. WILSON, C.C. BERG and D. FRENCH (eds). *Plenum: 1983. ISBN 0-306-41256-X. Np.*

Passive Solar Heating. J.R. WILLIAMS. *Ann Arbor: 1983. Pp. 304. ISBN 0-250-40601-2. £22.50.*

Biological Sciences

Bacterial Vaccines. RENE GERMANIER (ed.). *Academic: 1984. Pp. 441. ISBN 0-12-280880-0. \$65, £45.50.*

A Colour Atlas of Bacillus Species. J.M. PARRY, P.C.B. TURNBULL and J.R. GIBSON. *Wolfe Medical Publications: 1983. Pp. 272. ISBN 0-7234-0777-0. £25.*

Comparative Endocrinology. A. GORBMAN *et al.* *Wiley: 1983. Pp. 572. ISBN 0-471-06266-9. £33.20, \$46.50.*

Comparative Physiology of Sensory Systems. LIANA BOLIS, R. D. KEYNES and S.H.P. MADRELL (eds). *Cambridge University Press: 1984. Pp. 660. ISBN 0-521-25002-1. £65, \$99.50.*

The Effect of Ultraviolet Radiation on the Immune System: A Scientific Round Table. J.A. PARRISH (ed.). *Johnson & Johnson: 1983. Pp. 423. Pbk ISBN 0-931562-09-0. \$31.50.*

Fungus-Insect Relationships: Perspectives in Ecology and Evolution. QUENTIN WHEELER and MEREDITH BLACKWELL (eds). *Columbia University Press: 1984. Pp. 514. ISBN 0-231-0568-8. \$78.*

Genes, Organisms, Populations: Controversies over the Units of Selection. ROBERT N. BRANDON and RICHARD M. BURIAN (eds). *MIT Press: 1984. Pp. 329. ISBN 0-262-02205-2. £31.30.*

Herbivore Nutrition in the Subtropics and Tropics. F.M.C. GILCHRIST and R.I. MACKIE (eds). *Science Press: 1984. Pp. 779. ISBN 0-907997-03-1. Np.*

Hormonal Control of the Hypothalamo-Pituitary-Gonadal Axis. KENNETH W. MCKERNES and ZVI NAOR (eds). *Plenum: 1984. Pp. 622. ISBN 0-306-41800-2. Np.*

Human Alkaline Phosphatases. TORNGY STIGBRAND and WILLIAM H. FISHMAN (eds). *Alan R. Liss: 1984. Pp. 390. ISBN 0-8451-5016-2. £37.*

Human Genetics: An Introduction to the Principles of Heredity. By SAM SINGER. *W. H. Freeman: 1978. Pp. 139. Pbk ISBN 0-7167-0054-9. Pbk £10.95.*

IgA Mesangial Nephropathy. G. D'AMICO, L. MINETTI and C. PONTICELLI (eds). *Karger: 1984. Pp. 310. ISBN 3-8055-3877-4. DM 195, \$97.75.*

Immunological Aspects of Reproduction in Mammals. By D.B. CRIGHTON. *Butterworths: 1984. Pp. 529. ISBN 0-408-10865-7. £55.*

Immunoregulatory Processes in Experimental Allergic Encephalomyelitis and Multiple Sclerosis. By ARTHUR A. VANDENBARK and JEF C.M. RAUS. *Elsevier: 1984. Pp. 273. ISBN 7-0444-80570-2. \$75, Dfl. 195.*

Inclusion Compounds, Vol. 1 Structural Aspects of Inclusion Compounds Formed by Inorganic and Organometallic Host Lattices. J.L. ATWOOD, J.E.D. DAVIES and D.D. MacNICOL (eds). *Academic: 1984. Pp. 420. ISBN 0-12-067101-8. \$70, £48.*

Individual Susceptibility to Genotoxic Agents in the Human Population. FREDERICK J. DESERRES and RONALD W. PERO (eds). *Plenum: 1984. Pp. 518. ISBN 0-306-41679-4. \$69.50.*

Industrial Chemical Exposure: Guidelines for Biological Monitoring. R.R. LAUWERYS. *Biomedical Publications: 1983. Pp. 150. Pbk ISBN 0-931890-10-1. Np.*

Infection and Immunity in Farm Animals. R. PANDEY (ed.). *Karger: 1984. Pp. 255. ISBN 3-8055-3925-8. DM 196, \$98.25.*

International Review of Cytology. G.H. BOURNE and J.F. DANIELLI (eds). *Academic: 1984. Pp. 471. ISBN 0-12-364488-7. \$54, £42.*

International Review of Cytology, Vol. 86. G.H. BOURNE and J.F. DANIELLI (eds). *Academic: 1984. Pp. 331. ISBN 0-12-364486-0. \$45, £32.*

International Review of Experimental Pathology, Vol. 25. G.W. RICHTER and M.A. EPSTEIN (eds). *Academic: 1983. Pp. 391. ISBN 0-12-364925-0. \$56.*

The Interpretation of Ecological Data: A Primer on Classification and Ordination. By E.C. PIELOU. *Wiley: 1984. Pp. 263. ISBN 0-471-88950-4. £34.50.*

Ion Channels: Molecular and Physiological Aspects. WILFRED D. STEIN (ed.). *Academic: 1984. Pp. 394. ISBN 0-12-153321-2. \$75, £52.50.*

Issues and Reviews in Teratology, Vol. 1. H. KALTER (ed.). *Plenum: 1983. Pp. 354. ISBN 0-306-41239-X. Np.*

Osses and Reviews in Teratology, Vol. 2. HAROLD KALTER (ed.). *Plenum: 1984. Pp. 516. ISBN 0-306-41652-2. Np.*

Oxford Reviews of Reproductive Biology, Vol. 6. J.R. CLARKE (ed.). *Oxford University Press: 1984. Pp. 348. ISBN 0-19-857539. Np.*

Parasites and People. By D.W.T. CROMPTON. *Macmillan, London: 1984. Pp. 112. ISBN 0-333-32544-3. Hbk £10; pbk £4.95.*

Pathophysiology of Plasma Protein Metabolism. GUILIANO MARIANI (ed.). *Macmillan, London: 1984. Pp. 402. ISBN 0-333-34739-0. £40.*

The Pathway for Oxygen: Structure and Function in the Mammalian Respiratory System. By EWALD R. WEIBEL. *Harvard University Press: 1984. Pp. 425. Pbk ISBN 0-674-65790-X. Pbk £17.55.*

The Patient With End Stage Renal Disease, 2nd Edn. LARRY E. LANCASTER (ed.). *Wiley: 1984. Pp. 415. ISBN 0-471-88545-2. £28.50.*

Patient in the Womb. By E. PETER VOLPE. *Mercer University Press, Macon, GA, USA: 1984. Pp. 157. ISBN 0-86554-122-1. \$14.50.*

The Pea. By R. Kh. MAKASHEVA. *BALKEMA: 1984. Pp. 267. ISBN 90-6191-431-0. £31.50.*

Physiological Correlates of Human Behaviour, Vol. 2. Attention and Performance. ANTHONY GALE and JOHN A. EDWARDS (eds). *Academic: 1983. Pp. 245. ISBN 0-12-273902-7. \$39.50, £25.*

Physiologie Végétale. 1. Nutrition. By R. HELLER. *Masson: 1984. Pp. 244. Pbk ISBN 2-225-80326-9. Np.*

Physiology of Cell Aging. V.V. FROLKIS (ed.). *Karger: 1984. Pp. 208. ISBN 3-8055-3866-9. DM 196, \$98.25.*

Physiology of the Intestinal Circulation. A.P. SHEPHERD and D.N. GRANGER (eds). *Raven: 1984. Pp. 440. ISBN 0-88167-025-1. \$90.*

The Physiology and Pathophysiology of the Skin, Vol. 8. By A. JARRETT. *Academic: 1984. Pp. 2704. ISBN 0-12-380608-9. \$75, £49.*

Phytochemical Methods: A Guide to Modern Techniques of Plant Analysis, 2nd Edn. By J.B. HARBONE. *Chapman & Hall: 1984. Pp. 288. ISBN 0-412-25550-2. £17.50, \$33.*

Pig Diseases. By H.G. BELSCHNER. *Angus & Robertson: 1984. Pp. 152. ISBN 0-207-14487-7. £15.00.*

Plasma Protein Secretion by the Liver. H. GLAUMANN, T. PETER Jr and C. REDMAN (eds). *Academic: 1984. Pp. 497. ISBN 0-12-286-180-9. \$75, £45.*

Population Biology and Evolution. K. WOHRMANN and V. LOESCHKE (eds). *Springer-Verlag: 1984. Pp. 270. ISBN 3-540-13278-3. DM 112, \$40.80.*

Population Biology: Retrospect and Prospect. C.E. KING and P.S. DAWSON (eds). *Columbia University Press: 1983. Pp. 234. ISBN 0-231-05252-9. \$26.*

Primary Neural Substrates of Learning and Behavioral Change. DANIEL L. ALKON and JOSEPH FARLEY (eds). *Cambridge University Press: 1984. Pp. 385. ISBN 0-521-25472-8. £35.*

Protein Transport and Secretion. DALE L. OXENDER (ed.). *Alan R. Liss: 1984. Pp. 422. ISBN 0-8451-2614-8. £60.*

Recombinant DNA and Cell Proliferation. GARYS. STEIN and JANET L. STEIN (eds). *Academic: 1984. Pp. 360. ISBN 0-12-665080-2. \$49.50, £35.*

The Red Cell: Sixth Ann Arbor Conference. GEORGE J. BREWER (ed.). *Alan R. Liss: 1984. Pp. 608. ISBN 0-8451-5015-4. £66.*

Regression of Atherosclerotic Lesions: Experimental Studies and Observations in Humans. M. RENE MALINOW and VICTOR H. BLATON (eds). *Plenum: 1984. Pp. 355. ISBN 0-306-41732-4. Np.*

Reviews of Physiology, Biochemistry and Pharmacology, Vol. 97. R.H. ADRIAN *et al.* (eds). *Springer-Verlag: 1983. Pp. 176. ISBN 3-540-12135-8/0-387-12135-8. DM 88, \$38.*

The Revival of Injured Microbes. M.H.E. ANDREW and A.D. RUSSELL (eds). *Academic: 1984. Pp. 394. ISBN 0-12-058520-0. \$44, £29.50.*

Science and Gender: A Critique of Biology and its Theories on Women. By RUTH BLEIER. *Pergamon: 1984. Pp. 219. Hbk ISBN 0-08-030972-0; pbk ISBN 0-08-030971-2. Hbk \$25; pbk \$12.50.*

Seabird Energetics. G. CAUSEY WHITTOW and HERMANN RAHN (eds). *Plenum: 1984. Pp. 328. ISBN 0-306-41819-3. Np.*

Serengeti: Dynamics of an Ecosystem. A.R.E. SINCLAIR and M. NORTON-GRIFFITHS (eds). *University of Chicago Press: 1984. Pp. 389. ISBN 0-226-73640-7. Hbk £29.45; pbk £13.80.*

The Sex Life of Flowers. By BASTIAAN MEEUSE and SEAN MORRIS. *Facts on File: 1984. Pp. 152. ISBN 0-87196-907-6. \$19.95.*

Somatotensory Mechanisms. CURT VON EULER *et al.* (eds). *Macmillan, London: 1984. Pp. 396. ISBN 0-333-36943-2. £40.*

General

RSPB Nature Reserves. NICHOLAS HAMMOND (ed.). *Croom Helm: 1984. Pp. 189. ISBN 0-903138-11-5. £9.95.*

Shooting and the Countryside. By HUNTER ADAIR. *Oriel Press: 1984. Pp. 151. Hbk ISBN 0-85362-214-0; pbk ISBN 0-85362-215-9. Hbk £6.95; pbk £4.95.*

Theory and Measurement. By HENRY E. KYBURG Jr. *Cambridge University Press: 1984. Pp. 273. ISBN 0-521-24878-7. £22.50.*

The Total Synthesis of Natural Products, Vol. 6. JOHN APSIMON (ed.). *Wiley: 1984. Pp. 291. ISBN 0-471-09900-7. Np.*

The Touch of Midas: Science, Values and Environment in Islam and the West. By ZIAUDDIN SARDAR. *Manchester University Press: 1984. Pp. 253. ISBN 0-7190-0974-X. £20.*

Transition to Alternative Energy Systems: Entrepreneurs, New Technologies, and Social Change. THOMAS BAUMGARTNER and TOM R. BURNS (eds). *Bowker: 1984. Pp. 292. Pbk ISBN 0-86531-907-3. £21.75.*

Uncommon Sense: The Life and Thought of Ludwig von Bertalanffy, Father of General Systems Theory. By MARK DAVIDSON. *Houghton Mifflin: 1983. Pp. 247. Pbk ISBN 0-87477-334-2. Pbk \$8.95.*

Unlearned Language: New Studies in Xenoglossy. By IAN STEVENSON. *University of Virginia Press: 1984. Pp. 223. ISBN 0-8139-0994-5. \$17.50.*

Visual Masking: An Integrative Approach. By B.G. BREITMEYER. *Clarendon Press: 1984. Pp. 454. ISBN 0-19-852105-7. £25.00.*

Tests of influence

P.E. Bryant

Beyond IQ: A Triarchic Theory of Human Intelligence.

By Robert J. Sternberg.

Cambridge University Press: 1985. Pp.411. Hbk £25, \$39.50; pbk £8.95, \$14.95.

A PARADOX touches the study of intelligence. The subject is controversial — so controversial that it would be difficult to find even two psychologists who completely agree with each other on the nature of intelligence. Yet the psychologists' work in this area has had more impact outside the laboratory than anything else that has interested them, success which has merely increased the controversy. My own suspicion is that, if it were not for the obvious practical uses of these omnipresent tests, psychologists would have given up asking themselves about the underlying nature of intelligence as a far too abstruse and unrewarding question many years ago. But the very fact that the tests work, and that they seem to measure something which could reasonably be called intelligence, has posed a question which many have found impossible to resist investigating and trying to answer.

Robert Sternberg is the latest and almost certainly the most energetic in a long line of people who have devoted their time to attempting to say what intelligence means and what is responsible for intelligent behaviour. His approach is inclusive, and has become more so over the years. In this book, which is not his first on the subject, he sets out to combine three different approaches in what he calls a "triarchic" theory of intelligence. In order to understand why these three approaches have, on the whole, not been combined in the past, one needs to know something of the history of the study of intelligence.

It goes back to the first two pioneers of the intelligence test: Galton, whose tests were not a success, and Binet who, in contrast, produced the first effective intelligence test which has influenced every such test devised since then. The reasons for Galton's failure and Binet's success are important. Galton adopted a theoretical approach: he decided that intelligence consists of this and that ability and then produced tests to measure them. One test, for example, measured the speed of people's movements, another how well they named colours. These tests were certainly measures in the sense that different people produced different scores on them, but they did not measure intelligence. It soon turned out that they bore no relation to educational attainment, and that made it difficult to believe that they had anything to do with intellectual capacity.

Binet, too, had his theories, as Sternberg notes, but he did not let them play much part in his tests. His approach was prag-

matic, and he had one clear guideline. Whatever intelligence is, he argued, it increases during childhood. So he only accepted tests which older children were more likely to answer correctly than were younger ones. This approach worked. Binet's test did predict reasonably well how children would manage at school, and every other intelligence test since then has had a strong, though never perfect, relation to educational achievements.

But this practical success left an arresting theoretical vacuum. The tests themselves did little to reveal the underlying nature of intelligence. In the following 50 years or so the main attempts to fill this gap were made by people who used factor analysis as their main tool and their answers were distinctly uninspiring. They looked rather like the first part of a cooking recipe; intelligence was held to consist of this factor and this one and so on, and there was no sense of a system that actually worked. Sternberg gives reasonably short shrift to these theories. He is concerned with the workings of intelligence.

This is not at all surprising since Sternberg comes from a tradition in psychology which developed quite separately from the intelligence testing movement and which originally had nothing to do with individual differences. The study of information processing is concerned with the underlying processes involved in taking in, remembering and coordinating information about the environment, and its subject matter goes from the simplest kind of remembering or attention through to complex logical processes such as making sophisticated analogies or drawing inferences. Sternberg has always argued that this is the stuff of intelligence, and most of his work has taken the form of forging links between

studies of information processing and studies of differences in intelligence.

But now he thinks that this is not enough. His "triarchic" model is really a model of how to approach the question of intelligence rather than a fully fledged theory about intelligence. He argues that we need to know not only about the components of intelligence (the information processing involved) but also about the role of experience (particularly about the way that skills become automatic after a while) and, as well, about the context in which intelligent behaviour takes place. These three kinds of question are Sternberg's triarchy and the first half of the book is a convincing argument that all three are needed. In the second half he deals with particular problems which have always been closely associated with the study of intelligence, many of which — such as syllogisms and analogies — crop up a great deal in intelligence tests. His aim is to subject them to his triarchic approach, but here the analysis is often uneven though it is always very detailed. Sternberg's "componential" leanings take over. There is little in this part of the book about either context or the effects of experience. Sternberg has more research to do on these two questions, but given his phenomenal productivity there seems little doubt that he will manage it.

In the end the proof of Sternberg's pudding will be the same as it was with Galton's and Binet's. His scheme will only influence people outside the world of psychology, and to a great extent within it too, if it manages to produce good, new tests. Only then can we be sure that his scheme transcends the laboratory in a way that Galton's so signally did not. Sternberg seems to recognize this when at the end of the book he suggests several ways of improving tests of intelligence. He should try these out. In the meantime he has written an attractive and valuable account of the long and successful enterprise which began so well with Binet and still seems to be in very good hands. □

P.E. Bryant is Watts Professor of Psychology at the University of Oxford.

IMAGE
UNAVAILABLE
FOR
COPYRIGHT
REASONS

Upright education — the Geradhalter, an invention of a nineteenth-century educator, one Dr Schreiber, intended to improve children's posture by discouraging them from slumping forward. Schreiber used this and similar devices on his own children, one of whom was mentally deranged for most of his adult life. The illustration is taken from Robert Trivers's Social Evolution, published by Benjamin/Cummings.

On the rocks down under

Michael Audley-Charles

Phanerozoic Earth History of Australia.

Edited by J.J. Veevers.

Clarendon: 1985. Pp.418. £55, \$75.

REGIONAL geology is concerned with how particular areas of the outer part of the Earth have evolved. It attempts to account for all aspects of the rocks that can be mapped at and below the surface. The history of vertical and lateral movements, changes in climate, environments of deposition, rates and modes of erosion have to be interpreted from the great variety of observations, many of which are qualitative or descriptive. Magmatic, metamorphic, sedimentary and deformation processes have to be related in space and time to identifiable regional tectonic events that are often diachronous.

The elephant trap for writers of regional geology is to smother the reader in a snowstorm of detailed descriptions that fail to discriminate between the trivial and the vital. Readers thus afflicted, especially perhaps undergraduates, are led to associate regional geology with tedium. For this book, John Veevers has gathered together a team of writers who cover 575 million years of evolution of a region which represents a fortieth of the Earth's surface and a sixteenth of the continental lithosphere. Australia has no elephants and these authors eschew snow.

The apparent polar wander paths reveal that after some dalliance in the northern tropics during the early Palaeozoic, Australia has stayed in the Southern Hemisphere for most of the Phanerozoic. Records of oceanic palaeomagnetism for the last 160 Myr show how Australia's continental margins have been formed by a "can opener" type of rifting sequence which progressed from the north about 160 Myr anticlockwise around the margins to reach the Coral Sea about 60 Myr. These oceanic data do not reveal where or how these events began in what is now central New Guinea about 200 Myr. The origin of Australia's northern rifted margin before its late Cenozoic collision is not made explicit.

The evolving sequences of palaeoclimates and palaeoenvironments related to continental drift are dealt with deftly, the authors citing what they judge to be the three or four most sensitive and reliable diagnostic pieces of evidence for each Phanerozoic System of rocks, while a 25-page discussion of Australia's biogeography provides a glimpse of many unsolved problems and some of their seemingly more important implications. For example the tantalizingly uncertain data for Australian angiosperms leave us

not sure whether some flowering plants were able to enter northern Australia from Asia during the early Cenozoic nor whether the flowering plants were able to migrate from northern Australia into Asia in the late Mesozoic and early Cenozoic.

The main part of the text presents Australia's structural evolution through the Phanerozoic by recognizing only three morphotectonic elements: the northern convergent margin, the huge area of platform and the various divergent (rifted) margins on the west, south and east. The stratigraphic evolution is dealt with by the convincing device of recognizing three stratigraphic regimes: the Portoroo (Late Cretaceous and Cenozoic), Innamincka (Mid Carboniferous to Mid Cretaceous) and the Uluru (Cambrian to Mid Carboniferous).

This is not an encyclopaedic account of Australia's Phanerozoic geology. The reader should not expect to turn to the book for detailed description of particular

successions nor for a biostratigraphical discussion of the faunas and floras; the stratigraphy tends to be underlain by tectonic rather than biostratigraphic concepts. Sedimentological information is used throughout, but much of the text allows the fossils to sleep quietly (unmentioned) in their graves, so much so that some palaeontologists may turn in theirs. Perhaps that highlights the distance between stratigraphy and regional geology — a divorce not to be encouraged.

This book takes a refreshing approach to the synthesis of stratigraphy and tectonics, helped by writing which is concise without being over-condensed. Not only have Veevers and his colleagues produced an attractive and readable account of Australian geology, but more than in any other text I have seen succeed in conveying the excitement of regional geology. □

Michael Audley-Charles is Professor of Geology at University College London.

Chemists' enzymes

Richard Perham

Enzyme Chemistry: Impact and Applications.

Edited by Colin J. Suckling.

Chapman & Hall: 1984. Pp.255.

£17.65, \$36.

The Chemistry of Enzyme Action.

Edited by M. I. Page.

Elsevier: 1984. Pp.568.

\$66.50, Dfl. 179.

CHEMISTS have much to learn from, as well as contribute to, the study of enzyme catalysis. Although we no longer stand in awe at the catalytic power of enzymes, it is plain that there are still formidable problems to grapple with before we can claim a satisfactory understanding of their mechanisms. It is equally evident that we have not yet managed to apply all the lessons we have learned to, say, the industrial use of existing enzymes or the design of other catalysts.

Enzyme Chemistry is avowedly a didactic book, which sets out to document some of the past and to illuminate some of the future in the chemical study of enzymes and their products. As Kluger observes in an opening chapter which skilfully reviews the underlying physical organic chemistry of enzymes, if we were given the detailed three-dimensional structure of a protein of undisclosed catalytic specificity, it is unlikely that we could deduce the reaction catalysed. The goal of biomimetic chemistry — the design and construction of molecules that rival enzymes as catalysts — is thereby put in perspective. Shinkai concentrates on an encouraging approach to this goal that begins with a thorough study of the chemistry of coenzymes,

whereas Colin Suckling addresses the question of selectivity in chemical and enzyme synthesis. The fastidious selectivity and exquisite stereochemical preferences of enzymes are, of course, now well known but — as Suckling gently reminds us — their use has sometimes been treated, a little scornfully, as unacceptable in the perfect total synthesis. He cites some splendid examples of a more robust attitude and of its beneficial effect on biomimetic chemistry.

Industrial chemists have never shunned an enzyme or even whole cells as a means to an end. This theme is taken up by Hesp and Willard in a lively and optimistic discussion of enzyme chemistry and drug design. A rational approach to chemotherapy has seemed just around the corner for almost 50 years, ever since the discovery of the biochemical basis of sulphonamide antibiotics. That prospect still tantalizes but one perceives a quickening of step as the new sophistication of enzyme and protein chemistry and of molecular modelling begins to impinge. Similar optimism is advanced by Cane in a neat chapter on the enzymes responsible for the biosynthesis of natural products. Those chemists who still resist the use of enzymes in total synthesis will be shocked to find his advocacy of the use of recombinant DNA techniques to facilitate a study of the enzymes involved. Bornyl pyrophosphate synthetase, an enzyme of monoterpene biosynthesis, has an estimated turnover number of 10 mol of substrate per mol of enzyme per hour. Cane flatters this enzyme (he is obviously but justifiably fond of the pathway) by referring to its catalytic power as extremely modest. Given that this may be typical of secondary metabolic enzymes, chemists would be foolish indeed to reject the help that molecular biology can offer them. They might even indulge in the delights of site-directed mutagenesis.

The book is completed by an article (Brown and Smith) on metal ions in biological systems—inorganic chemistry is not neglected—and an assessment by Keith Suckling of the effect of enzymological ideas on seemingly unrelated areas of biochemistry.

For a multi-authored book, the style and standard of presentation is unusually even. The authors communicate their affection for the subject and display a refreshing awareness of the benefits to chemistry that flow from enthusiastically embracing the study of enzymes, as well as the metabolites they convert. A significant criticism is that the references terminate in mid 1982 so that, for example, much of the current achievement of the molecular biological approach, let alone its potential, goes unrecorded. Biology is without doubt well placed to repay chemistry.

The Chemistry of Enzyme Action is part of the series entitled *New Comprehensive Biochemistry*. It is a large book (15 chapters) and includes most of the ground covered by *Enzyme Chemistry*, but goes well beyond it both in the range of topics and the depth in which they are treated. The principal aim of the book is to describe the physical and organic basis of enzyme action. Because the approach is mechanistic, the opening chapters concentrate on the energetics and substrate interactions of enzymes (Page, Kollman, Williams) and the use of kinetic (Engel, Douglas and Wilson) and isotope (Williams) methods in exploring mechanisms. Several systems are then selected for a detailed chemical scrutiny: enzyme reactions that utilize imine formation (Hupe), enzymes that require pyridoxal phosphate (Akhtar, Emery and Robinson), folate and bioprotein cofactors (Benkovic and Lazarus), glycosyl transfer (Sinnott) and vitamin B₁₂ (Brown). Biomimetic systems are represented in the concluding chapters on micelles and similar aggregates (Brunton), cyclodextrins (Komiya and Bender) and crown ethers (Stoddart).

Despite its size, the book therefore remains selective in its coverage. The complementary part of this story, the participation of the enzyme protein and its structure in the enzymic reaction, is consciously left very much on one side. Thus, individual chapters will be read with profit by specialists but I fancy that the average undergraduate would find the rigour and detail too demanding. The logical place for such a book is the departmental library and the shelves of enzymologists with relevant interests. Typographical errors, while they do not abound, pop up from time to time, but the overall presentation is not unpleasing. All the authors can be congratulated on the comprehensive reference lists they have compiled. □

Richard Perham is a Reader in the Department of Biochemistry, University of Cambridge, and President of St John's College, Cambridge.

An independent view

John H. Lawton

The Ecological Web: More on the Distribution and Abundance of Animals.

By H.G. Andrewartha and L.C. Birch.
University of Chicago Press: 1985.
Pp. 506. \$40.25, £33.25.

SCIENCE is most exciting when two alternative views of the world clash. The battle in population dynamics is between those who believe that density-dependent processes play a significant part in determining the distribution and abundance of animals, and those who do not. Andrewartha and Birch have led the density-independent school, ever since their book *The Distribution and Abundance of Animals* appeared in 1954. Their views have not changed significantly over the past 30 years, although the arguments have become more sophisticated and the controversy has recently enjoyed new vigour.

The central theme of *The Ecological Web* is a "theory of environment" based on "envirograms" — dendrograms whose branches trace pathways of influence impinging upon study-species. There are four directly acting components, resources, mates, malentities (that is, hazards) and predators, all of whose actions can be influenced by a cascade of steps in other parts of the envirogram. Andrewartha and Birch's views of population dynamics are therefore particular and detailed. They dismiss general, analytical models in two sentences: "We cannot see many testable hypotheses emerging from such abstract models. . . . So we say no more about them" (p.187). Approximately the first two-thirds of the book is devoted to describing and establishing their theory of environment; there follow in the final third several case-histories, on rabbits, two birds, marine fish (briefly), three species of insects and man. Only the spruce budworm, fish and man are not Australian examples.

I ought now to make my own position clear. Unlike Andrewartha and Birch, I find abstract general models of ecological processes useful, and I do think that density-dependent processes are largely irrelevant in understanding population dynamics. However I set out to read *The Ecological Web* with an open mind, but in the end found it lacking in many ways.

The theory of environment is largely verbal and pictorial, but is used to attack more formal, mathematical theory. The result is rather like comparing apples and oranges. The weakness of predominantly verbal arguments is best illustrated by an example. In the section "Competitive Coexistence and Displacement" the authors conclude, after reviewing several case-histories, that species can live together "because they are alike — not identical Conversely, when two species fail

to coexist it is usually because they differ with respect to their success in professing a niche that they have in common". They seem not to notice that this conclusion is tautological; nor does it take us any nearer to a general understanding of competitive coexistence or exclusion.

Perhaps more seriously, Andrewartha and Birch often do not seem to understand many of the theories they either attack, or so lightly dismiss as futile. Moreover, they too often pass over evidence that contradicts their view of the world. Consider, for example, their efforts to explain population cycles in animals such as snowshoe hares or lemmings. They use neutrally stable Lotka-Volterra models (p.255) yet claim that "oscillations decline and finally disappear"; they do not. More interesting and realistic general models are ignored, as are efforts to test them^{1,2}. Alternatively, take cases of "apparent competition" via shared enemies, or the role of spatial heterogeneity in population dynamics. They have much to say about both, but all their points are covered in more rigorous ways by existing general theoretical models. A small number of other examples will suffice to make my point. Great tit populations³ are not limited by nest sites, as claimed. Population density does have important effects on the dynamics of tsetse flies⁴, spruce budworms⁵ and carabid beetles^{6,7}. Sadly, I have to conclude that Andrewartha and Birch often develop their arguments by ignoring counter-evidence.

It appears that most of the confusion in this area of ecology centres on whether or not people are more impressed by population fluctuations, or by the fact that, despite such fluctuations, numbers of many species remain within comparatively narrow limits. Clearly not all populations are well regulated, but many are. It is important to know whether there are systematic patterns in the relative contributions of density-dependent and density-independent processes to the population dynamics of species from different taxa and habitats; but you will not find the answer to such questions in this book. Much has happened in population biology in the 30 years since the publication of *The Distribution and Abundance of Animals*. Because its treatment of these developments is so inadequate, *The Ecological Web* does little to advance the case for density independence. □

John H. Lawton is a Reader in the Department of Biology, University of York.

- Peterson, R.O., Page, R.E. & Dodge, K.M. *Science* 224, 1350-1352 (1984).
- Bryant, J.P., Chapin, F.S. & Klein, D.R. *Oikos* 40, 357-368 (1983).
- Perrins, C.M. *British Tits* (Collins, London, 1979).
- Rogers, D.J. & Randolph, S.E. *A. Rev. Ent.* 30, 197-216 (1985).
- Royama, T. *Ecol. Monogr.* 54, 429-462 (1984).
- Baars, M.A. & Van Dijk, T.S. *J. Anim. Ecol.* 53, 389-400 (1984).
- Brunsting, A.M.H. & Heessen, H.J.L. *J. Anim. Ecol.* 53, 751-760 (1984).

Classified material

V.H. Heywood

Guide to Standard Floras of the World.
By D.G. Frodin.
Cambridge University Press: 1985.
Pp.619. £95, \$175.

THE best description of this remarkable compendium is given by the subtitle, "An annotated, geographically arranged systematic bibliography of the principal floras, enumerations, checklists, and chorological atlases of different areas". The second part of the book, some 500 pages, comprises just that. There are, however, really two separate elements in this enterprise — a thoughtful analysis of *Floras* and floristics, and the application of this analysis to the geographically arranged guide to the most useful *Floras* and checklists of vascular plants of the world, selected according to the concept of "standard" works. The idea for this compilation dates back to 1962, and we are told that the main catalyst was the publication of the so-called "Green Book" produced by the *Flora Europaea* editorial committee for the guidance of contributors, in which a list of "standard" *Floras* of Europe was given.

Preparation of the *Guide* began in 1970 and was completed in 1975. It is a tribute to the author that he was able to contemplate such a work based in Papua New Guinea, scarcely one of the world's leading bibliographical centres (in a remarkable piece of understatement he comments that the "remoteness of Port Moresby was ... a handicap"!). Publication has been delayed for nearly six years after the submission of the manuscript to the publishers, and with a cut-off year of 1980 the text is already seriously out of date in many places.

Floristic projects are advancing so rapidly today that the whole question of maintaining an up-to-date database needs careful consideration. It is not at all clear to me that a hard-copy guide such as this, however selective it is, is the most effective instrument for bibliographical purposes. The author explains, however, that his aim was not just to present a classified bibliography but, by being analytical and interpretative, to act as a mirror on the progress of the subject. His ideas are expanded in the first section of the book. This includes essays on the style of *Floras* which are an important contribution to the literature on the subject. Frodin believes that the basic role of a *Flora* should still be a utilitarian and practical one — "inventory, identification and essential related data". He further echoes the increasingly expressed view that, for tropical countries in particular, floristic works should be written in such a way as to have a wide public appeal.

A disappointing feature of this introductory section is that although the author looks to the future, and emphasizes the

need for a reappraisal of the ways in which floristic data are stored and presented, he does not put forward any very concrete suggestions. It is unfortunate that the book was completed just at the time when several new approaches to floristics were being implemented, most of them related to electronic data-handling. Thus a whole series of developments in the use of databases in systematics are omitted or only touched upon, such as the Vicieae database at Southampton, the revived Flora North America project (which will be heavily dependent on database technology) and the European Science Foundation Taxonomic Information System and database at Reading, not to mention the substantial reorganization of taxonomic services, involving electronic data processing, that several major taxonomic institutions are currently engaged upon.

In the main part of the *Guide*, Frodin's essays and annotations make interesting reading and present a very idiosyncratic interpretation of the facts reported. This material would have been better published separately. Indeed, altogether it is highly questionable whether it was wise for a world bibliography of *Floras* to have been undertaken as a one-man operation. One cannot but admire the author's enormous achievement in compiling this unique, expensive and fascinating work, but, at the same time, one cannot help feeling that the time for personal statements in this field has been bypassed by events, as Frodin himself hints. □

V.H. Heywood is Professor in the Department of Botany, University of Reading.

Past futures

D.Q. Bowen

Late Quaternary Environments of the Soviet Union.

Edited by A.A. Velichko.
University of Minnesota Press/Longman:
1985. Pp.327. \$45, £50.

This is a third, companion volume to *Late Quaternary Environments of the United States* (reviewed in *Nature* 311, 488; 1984), and similarly arises from the USA-USSR bilateral agreement on the environment and meetings of the scientists sponsored by that agreement. This English language edition brings to the West a whole new range of data, new not least because much of the research summarized here is the product of recent years only.

As such, the book is self-evidently a publication of some significance. Given the vast area of the Soviet Union (exceeding that of the United States and Canada combined), and the diversity of the environments it encompasses, it is inevitable that the quality of coverage will be uneven. But slowly the long-standing, mostly geo-

morphologically based work is being supplemented by sedimentological, chronological and biostratigraphical studies. Appropriately enough, nearly all those Soviet scientists who have worked so hard to transform the situation figure in the book.

There are 30 chapters, many of which cover topics of international interest: the Late Pleistocene glaciation of the northern Soviet Union (in which a carefully argued case that large ice-sheets *did not* occupy the sites of the Barents and Kara Seas is prominent); mountain glaciation; permafrost in the Late Pleistocene and Holocene; and loesses, with a not unexpected demonstration, which nevertheless fills another gap in global correlation, that nine soil complexes are found above the Matuyama-Brunhes reversal in Central Asia. Furthermore, in the Tadzhik Depression the Olduvai reversal is identified; thus parallels emerge between Europe and points farther east, and ultimately with the global deep-sea isotope signal. An account of vegetational history gives a welcome insight into a greatly expanded field of endeavour, even though the data are regionally uneven and of varying quality, while a fascinating chapter on inland sea basins shows how sea-level change varies according to global as well as regional controls: thus 60,000 years ago the Caspian Sea rose to 76m above its present level, an event attributed to reduced evaporation and increased input from the Volga.

Palaeoclimatic reconstruction has long been a strength of Soviet scientists and included are convincing reconstructions of the principal stages along with inferences on their atmospheric circulation. It is shown that the "glacial monsoon" of the western Soviet Union, created by juxtaposition of cold high-pressure zones over the ice sheet and warmer low-pressure systems farther south, played a vital role in the transport and deposition of the loess deposits so critical for correlation.

Many readers will repair eagerly to the section on the dispersal of primitive cultures, not least in the hope that evidence from Siberia will throw light on the timing of the entry of man into the New World via the Bering Strait and on the attendant possible extinction of the megafauna. Unfortunately, however, the evidence from many late Palaeolithic sites in the Soviet Union is equivocal and contributes little to resolution of this problem.

The introduction to the English edition by H.E. Wright and C.W. Barnosky pulls the wide range of chapters together in an integrative and informative manner. Given the continuing exploration and development of the natural resources of the Soviet Union, it seems likely that much of the Late Quaternary framework evident there will be further elaborated and be of still greater benefit to Quaternary scientists elsewhere.

D.Q. Bowen is Professor of Physical Geography at The University College of Wales, Aberystwyth.

Eye movements recorded

This month's sampling of new products includes a sensitive system for recording human eye movements and a bench-top dry ice machine.

● Microtome manufacturers Temtool Ltd claim that a few years ago disposable 'razor' blades dominated the market, but now the economics of disposable versus re-usable knives is such that the re-usable knife is again competitive, especially when an automatic knife sharpener is used. The Rotosharp 2 from Temtool is an automatic sharpener that takes a wide variety of knife sizes and types (including tungsten carbide) and produces good edges in a few minutes. *Reader Service No. 100.*



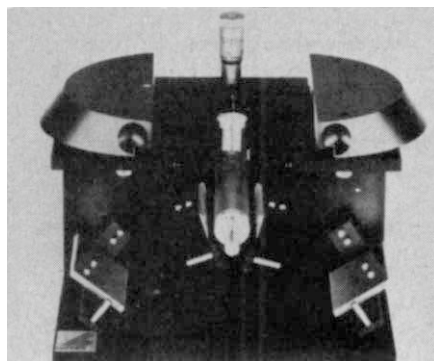
A 0.1-micrometre line pattern on silicon, written on a 0.35-micrometre-thick PMMA resist with the Philips Beamwriter.

● Sophisticated software, operating on a simple question-and-answer basis, has virtually eliminated manual adjustment of the Beamwriter EBPB-4 vector scan lithography system from Philips. Allowing line-widths down to 0.1 micrometres, the EBPB-4 offers true sub-micrometre resolution. Developed from the world's first commercial production-oriented vector scan electron beam pattern generator (EBPG) system, the EBPB-4 Beamwriter has been specially designed for direct writing onto silicon or gallium arsenide wafers but can also be used for making masks and reticles. *Reader Service No. 101.*

● For applications in the laboratory where only the standard, 13A a.c. mains power supply is available, Lenton Thermal Designs Ltd has created a range of programmable bench-mounted furnaces: the Ultraspeed and the ECF series. The Ultraspeed reaches maximum operating temperature in only 10 min, and is available in two chamber sizes, 1.5 litres and 2.0 litres. ECF furnaces are available as a general purpose laboratory furnace and as an ashing furnace for the ashing of foods prior to microanalysis. *Reader Service No. 102.*

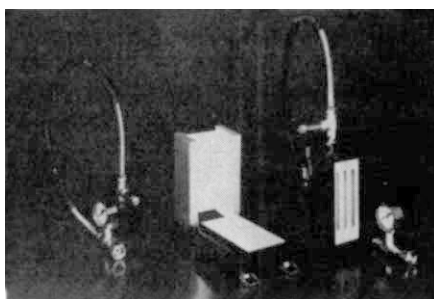
These notes are based on information provided by the manufacturers. Use the reader service card bound inside the journal to obtain further information.

● Now available from Barnes Analytical/Spectra-Tech is the Collector II, an advanced version of their FTIR accessory for DRIFT (diffuse reflectance infrared Fourier transform) spectroscopy. A major development is the Blocker, a device which 'blocks' and greatly reduces or eliminates the specular reflectance component. The Collector II is especially suited for obtaining spectra from fibres, powders, foams, paint chips, catalysts (zeolites) and coal, as well as LC and HPLC eluents. *Reader Service No. 103.*



Collector II for DRIFT spectroscopy.

● The SP8773XR absorbance detector, designed for routine operation at low detection limits, complements the Spectra-Physics line of modular, extended-range LC systems. LED display indicates real-time status, and provides self-diagnosis data for easy troubleshooting. Precision flow cells are readily interchanged to accommodate micro to preparative analyses. *Reader Service No. 104.*



Bench-top dry ice from Polyfoam Packers.

● The Insta Ice Machine, available from Polyfoam Packers Corporation, is a portable bench-top dry ice machine weighing only six pounds. Blocks can be made in 1 lb or 2 lb sizes that do not require any further compacting. No electricity or batteries are needed — all that is required is a cylinder of liquid CO₂ with a syphon. *Reader Service No. 105.*



Skalar's device for measuring eye movement.

● Precise recording of human eye movements is possible with a new system from Skalar of Holland. It involves a scleral induction coil, embedded in a flexible annulus of silicone rubber, which adheres tightly to the limbus of adult human eyes, concentric with the cornea, in a magnetic field. Two a.c. magnetic fields of similar frequency but in spatial and phase quadrature are created by field coils. An a.c. potential is induced in a sensor coil mounted on the eye and the magnitude and phase of the signal depend on the orientation of the eye in the field. *Reader Service No. 106.*

● Anti-tubulin, useful for localizing microtubules in cells, quantitating the amount of tubulin in cells and tissue by radioimmunoassay and for gene cloning and studies of transformed cells, is available from Polysciences, Inc. The antibody can be used for immunocytochemical staining at dilutions of 1:50 to 1:100. *Reader Service No. 107.*

Advertisement

LIPOSOMES NEW!

reproducible — homogeneous
— extremely rapid —

The new commercial LIPOSOMAT® provides reproducible preparation of uniformly sized unilamellar liposomes (formation within 10 minutes!) with very high encapsulation efficiency for both hydrophilic and/or lipophilic agents. Liposome size selectable between 25 and 600 nm diameter. Lipid loss far below 1%. Sample volume normally 5-10 ml or up to 200 ml if used in combination with an additional apparatus.

The LIPOSOMAT® is used especially for studying drug/protein/cell — liposome interactions in vitro and in vivo, for reconstitution experiments, and for the transfer of genetic material in gene- and biotechnology.

DIANORM-Geräte, POB 126, D-8000 MUNICH 65, FRG.

Reader Service No. 6.

Women in science and engineering

from Richard Pearson

One of the biggest changes in the employment scene in the second half of this century has been the rising participation of women in the labour market.

IN the 1930s only one in ten women of working age in the United Kingdom were in employment, now nearly half are in or seeking employment; and women now account for 43 per cent of the labour force.

To date, however, women have been concentrated in the unskilled, lower-paid jobs, often working part-time hours. In terms of paper qualification, far fewer women are qualified than men. The 1981 *Census of Population* shows that while 7 per cent of adult men were graduates, the proportion of women so qualified was less than half this figure, at 3 per cent. However, the proportion of women gaining degrees has been rising fast towards equality, so that in 1983 42 per cent of those in higher education in 1983 were women compared with under 30 per cent twenty years earlier. Although more women are becoming qualified, the range of qualifications they are gaining is still markedly different from their male counterparts.

One of the most critical points for career choice is in the school system, where, by the age of 13, most girls have committed themselves to arts rather than science subjects, which severely constrains their later entry into science-based careers and courses in higher education. In terms of school qualifications, the proportion of girls gaining two or more A level qualifications in science, the usual minimum entry qualification for science or engineering courses in higher education, has barely changed, averaging about one in five.

Thus although the proportion of women going into higher education has risen dramatically over recent years, and they now account for more than half the intake into many medical schools, the changes on science and engineering courses have been far less marked. While in the case of engineering the proportion has more than doubled, women still account for only one in ten students, and the proportion of women on science courses has risen only slightly (Table 1). Of the women graduating

last year, nearly two-thirds had art or social science degrees, only 21 per cent had science degrees and a mere 4 per cent had engineering degrees.

Curiously, girls' preferences within engineering do not seem to conform to the expected pattern in relation to individual subjects, thus women represent an above average 12 per cent of the total in chemical engineering, but under 6 per cent in electrical engineering and electronics, a field generally regarded as having a less traditional heavy engineering image. This rather lower representation may be because the competition for places is much harder in these subjects.

Table 2 Percentage of women in main graduate occupations.

Teachers	56.0%
Medics	29.6%
Scientists	16.0%
Accountants	8.1%
Lawyers	15.1%
Architects	4.2%
Engineers	1.2%

Source: *Census of Population*, for 1981.

In terms of jobs, the contrasts with men are even more marked. The most fundamental difference in 1981 was that almost a third of the women graduates (about 200,000) were not working or seeking work, mostly presumably for family reasons, while among men the corresponding figure was only 11 per cent. Interestingly, similar proportions of men and women graduates were unemployed — three per cent. Of the 450,000 working women, almost half were in teaching and indeed in this occupation they outnumbered their male counterparts. In medicine, women are relatively well represented, accounting for nearly one in three of all medical and dental practitioners. However, only about one in six of scientists are women, and fewer than one in eighty of the engineers (Table 2). There are far more graduate women in accountancy and in law than in engineering.

To remedy this low representation of women in science and engineering, for which both women and society generally suffer, the year 1984 was designated as WISE — Women in Science and Engineering — Year. Sponsored by the Engineering Council and the Equal Opportunities Commission, the main aim of the campaign was to educate and encourage girls,

through a wide range of promotional activities, to take up careers in science and engineering. Even before WISE Year there have been initiatives aimed at encouraging schoolgirls to consider engineering careers. These have included special "Insight" programmes to allow sixth-formers to have close experience of engineering over a period of a weeks in industry, special scholarships and local school-industry liaison activities. The latter have not always been a success, however, with one girls' school returning a package of engineering careers leaflets with the comment "This must have been sent to us in error, this is a girls' school".

WISE Year included over 500 special conferences and events in schools, colleges and employers' premises, and many activities are continuing into 1985 and beyond. A limited number of new scholarships for girls was awarded during 1984 by the Engineering Training Board and others, but because the year was primarily aimed at raising awareness, it will be several years before any effective evaluation can be made, when any changes in subject choice among girls in schools and higher education will start to become apparent. At any rate, the organizers are already claiming that the year was a success.

An added urgency for women to enter technology-based courses was the fact that 1982 saw the number of 18-year-olds in the population, the traditional entry group into higher education, start to fall. The number will decrease from 1,000,000 to 650,000 in 1995 with most of the fall occurring during the latter part of the period. While the implications of this fall in terms of demand for higher education places have been hotly debated for several years now, in the cases of less popular subjects such as engineering, such a fall could have a significant effect on the numbers graduating. A rising relative demand by women will help mitigate against the overall fall and an increasing representation by women on engineering and technology courses may be the only means of ensuring an adequate future supply of graduates with key skills. Thus meeting the needs of the market and providing equality of opportunity could well become self-supporting policies in the late 1980s. □

Table 1 Women first-year students as % of all students

Medicine	49.0
Engineering and technology	9.5
Science	33.1
Social science	44.4
Language, literature	69.0
Other arts	54.6
All subjects	40.7

Source: UCCA figures for 1982.

Richard Pearson in at the Institute of Manpower Studies, Mantell Building, University of Sussex, Brighton BN1 9RF, UK.